

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2002, 17:26:02 ; Search time 35 Seconds  
(without alignments)  
2125.229 Million cell updates/sec

Title: US-09-839-894-10

Perfect score: 1886

Sequence: 1 MNKILFIFFLFFSSVLFTEFA.....EAGQYMGINVTFTPSQTL 361

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: SP archaea: \*  
2: SP bacteria: \*  
3: SP fungi: \*  
4: SP human: \*  
5: SP invertebrate: \*  
6: SP mammal: \*  
7: SP mnc: \*  
8: SP organelle: \*  
9: SP phage: \*  
10: SP plant: \*  
11: SP rodent: \*  
12: SP virus: \*  
13: SP vertebrate: \*  
14: SP unclassified: \*  
15: SP virus: \*  
16: SP bacteriophage: \*  
17: SP archaea: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1886	100.0	361	2	Q93667
2	949.5	50.3	363	2	Q47115
3	930.5	49.3	364	2	Q47119
4	275.5	14.6	359	16	Q9XDS1
5	127	6.7	1881	16	Q8RCK2
6	115.5	6.1	700	2	Q8VUS0
7	115.5	6.1	847	2	P71132
8	111.5	6.0	865	15	Q8Q7H7
9	111.5	5.9	743	16	Q8XNM2
10	110	5.8	846	2	P71133
11	109.5	5.8	2454	5	Q8R2G3
12	109	5.8	807	5	Q8R7V5
13	108	5.7	872	5	Q8R7V5
14	108	5.7	2529	16	Q25579
15	107	5.7	304	16	Q8XAX2
16	105.5	5.6	430	3	Q96VN2

17	105.5	5.6	843	16	Q8YKQ8	Q8YKQ8 anabaena sp
18	105	5.6	598	16	Q8Y717	Q8Y717 listeria mo
19	104.5	5.5	1441	16	Q9CFL1	Q9CFL1 lactococcus
20	104	5.5	803	10	Q9AUS5	Q9AUS5 oryza sativ
21	103.5	5.5	691	16	Q9RZ57	Q9RZ57 delinococcus
22	103.5	5.5	1269	17	Q97206	Q97206 sulfolobus
23	103.5	5.5	2399	16	Q9ZK59	Q9ZK59 helicobacte
24	103	5.5	1754	16	Q8R8D1	Q8R8D1 thermoaer
25	102.5	5.4	1736	5	Q95PH7	Q95PH7 dictyostell
26	102.5	5.4	2747	2	Q91800	Q91800 aeromonas s
27	102	5.4	692	17	Q97YM6	Q97YM6 sulfolobus
28	101	5.4	1238	16	Q8ZJ21	Q8ZJ21 yersinia pe
29	100	5.3	1014	2	Q46962	Q46962 escherichia
30	100	5.3	1752	10	Q9LX60	Q9LX60 arabidopsis
31	99.5	5.3	381	16	Q8R8X2	Q8R8X2 thermoaer
32	99.5	5.3	509	5	Q8R7S0	Q8R7S0 drosophila
33	99.5	5.3	1300	16	Q9CKZ7	Q9CKZ7 pasteurella
34	99.5	5.3	5627	16	Q91120	Q91120 pseudomonas
35	99	5.2	456	2	Q48508	Q48508 lactobacill
36	99	5.2	802	16	Q92C16	Q92C16 listeria in
37	99	5.2	839	2	P77792	P77792 chlamydia p
38	99	5.2	836	16	Q9PHY4	Q9PHY4 campylobact
39	99	5.2	1355	16	Q8XXK6	Q8XXK6 clostridium
40	98.5	5.2	793	5	Q8SSR7	Q8SSR7 dictyostell
41	98.5	5.2	4919	2	Q9ZHL0	Q9ZHL0 haemophilus
42	98	5.2	587	16	Q92FA2	Q92FA2 listeria in
43	98	5.2	1014	2	Q47106	Q47106 escherichia
44	97.5	5.2	855	15	Q6961	Q6961 feline immu
45	97.5	5.2	1755	17	Q26812	Q26812 methanobact

## ALIGNMENTS

RESULT 1  
ID Q93667 PRELIMINARY: PRT: 361 AA.  
AC Q93667:  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE Minor pilin protein CsaE.  
GN CsaE.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=EL1881A;  
RA Altboun Z.D., Levine M.M., Galen J.E., Barry E.M.;  
RT "Isolation and characterization of EPEC CS4 fimbriae encoding genes,  
RT and their expression in Shigella flexneri 2a guaba strain CVD 1204.";  
RL Submitted (Aug-2000) to the EMBL/Genbank/DBJ databases.  
DR EMBL: AF296132; AAK97137.1; -  
SQ SEQUENCE 361 AA: 40102 MW: 2E1E74ABDB0EB353 CRC64:

Query Match 100.0%; Score 1886; DB 2; Length 361;  
Best Local Similarity 100.0%; Pred. No. 8e-137;  
Matches 361; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MNKILFIFFLFFSSVLFTEFAVSADKIPGDESTITFGPRDRNESSPKHILNNHTAYSE 60  
DB 1 MNKILFIFFLFFSSVLFTEFAVSADKIPGDESTITITFGPRDRNESSPKHILNNHTAYSE 60  
QY 61 SHTLVDKMTFLILSSHTNLNGACPTSENSSSSVSGETNITQFTEKRSILIRELOIKCY 120  
DB 61 SHTLVDKMTFLILSSHTNLNGACPTSENSSSSVSGETNITQFTEKRSILIRELOIKCY 120  
QY 121 KOLLKFSVNCPSGLFTLNSAHFNCNKAASGASLYIYPAGELKNLPFGGIWDTATLKYK 180  
DB 121 KOLLKFSVNCPSGLFTLNSAHFNCNKAASGASLYIYPAGELKNLPFGGIWDTATLKYK 180

```

OY 181 RRYSEYGYTYITITKLDKNGIQIWLPOFKSDARVDLMLRPTGGGTYIGRNSVDMCFY 240
    |||||
DB 181 RRYSEYGYTYITITKLDKNGIQIWLPOFKSDARVDLMLRPTGGGTYIGRNSVDMCFY 240
OY 241 DGYSTNSSSLEIFPODNNPKSDGKFLRKINDOTKEIAYTLLSLAGKSLTPNGTSLNT 300
    |||||
DB 241 DGYSTNSSSLEIFPODNNPKSDGKFLRKINDOTKEIAYTLLSLAGKSLTPNGTSLNT 300
OY 301 ADAASLETNNRITAVTAPPEISVPLVLCMPGRLQDADAKVENPEAGOVGNINVTFTPSQOT 360
    |||||
DB 301 ADAASLETNNRITAVTAPPEISVPLVLCMPGRLQDADAKVENPEAGOVGNINVTFTPSQOT 360
OY 361 L 361
DB 361 L 361

RESULT 2
O47115 PRELIMINARY; PRT; 363 AA.
ID 047115;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE COOD precursor.
GN COOD.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
NCBI_TaxID=562;
RX MEDLINE=94344028; PubMed=7915003;
RA Froehlich B., Karakashian A., Melsen L.R., Wakefield J.C., Scott J.R.;
RT "Cooc and Cood are required for assembly of CSI pil1.";
MOL. Microbiol. 12:387-401(1994).
EMBL: X76908; CAAS4230.1; -.
DR InterPro: IPR001412; TRNA-synL.
DR PROSITE: PS00178; AA_TRNA_LIGASE_I; UNKNOWN_1.
KW Signal.
FT SIGNAL. 1 18 POTENTIAL.
FT CHAIN 19 363 COOD.
SQ SEQUENCE 363 AA; 40139 MW; AA948BADEDE5F72 CRC64;

Query Match
Best local Similarity 50.3%; Score 949.5; DB 2; Length 363;
Matches 197; Conservative 55; Mismatches 98; Indels 21; Gaps 11;

3 KILFIETLFFSSVLTFAVSADKIPGDE--STINIF-GPR-DRNESSPKHILNNHITAY 58
    |||||
2 KIIFFIFL---SIIFSAVVSAGRPETTVGNLTKFSQAPRLDRSVQSPIYINIFNHVAGY 57
OY 59 SESHTLYDRMTFLCLSHNTLNGACPTSENPSSSVS-GETNITLQTEKRSILKRELQI 117
    |||||
DB 58 SLISLSDRLVPLCTSSSNVNGACPTI---GTSVOYGTFTTTLQTEKRSILKRNINL 114
OY 118 KGYKOLLFKSVNCPG--LTLNSAHFNCNKNA-ASGASLYLIPAGELKNLPGGINDAT 174
    |||||
DB 115 AGKKKPIWENQSCDFSLMVLNKSMSGCAHGANGLMLLYIPAGINKLPGGIMEAT 174
OY 175 LKLRVRRYSET---YGYTYINITIKLTDKNGIQIWLPOFKSDARVDLMLRPTGGGTYI 230
    |||||
DB 175 LILRLS-RYGEVSTHGNVTYITVDLTDKNGIQIWLPGFHSNPRVDLMLRPGNKYS 233
OY 231 GRNSVDMCFYDGYSTNSSSLEIFPODNNPKSDGKFLRKINDOTKEIAYTLLSLAGKSL 290
    |||||
DB 234 GSNSLDMCFYDGYSTNSDSVNIKFQDNNPTNSSEYNYLYKG-GEKLPYAVSLMGKIF 292
OY 291 TPNNGTSLNADASLETNNRITAVTAPPEISVPLVLCMPGRLQDADAKVENPEAGOVGNIN 350
    |||||
DB 293 YFVNGOSFTINDSSVLETNNRITAVTAPPEISVPLVLCMPARLLINADYVNPADAGQISQI 352
OY 351 NVTFPSSQTL 361

```

```

DB 353 YITFTPSVENTL 363
    :|||:
OY 353 YITFTPSVENTL 363

RESULT 3
O47119 PRELIMINARY; PRT; 364 AA.
ID 047119;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE COOD precursor.
GN COOD.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
NCBI_TaxID=562;
RX MEDLINE=96071908; PubMed=7591145;
RA Froehlich B.J., Karakashian A., Sakellaris H., Scott J.R.;
RT "Genes for CS2 pil1 of enterotoxigenic Escherichia coli and their
interchangeability with those for CSI pil1.";
MOL. Infect. Immun. 63:4849-4856(1995).
EMBL: Z47800; CAAS7763.1; -.
KW Signal.
FT SIGNAL. 1 18 POTENTIAL.
FT CHAIN 19 364 COOD.
SQ SEQUENCE 364 AA; 40979 MW; EC7AC738ADE195B6 CRC64;

Query Match
Best local Similarity 49.3%; Score 930.5; DB 2; Length 364;
Matches 177; Conservative 64; Mismatches 118; Indels 7; Gaps 3;

1 MKKILFIETLFFSSVLTFAVSADKIPGDESTINIFGPRDRN--ESSPKHILNNHITAY 58
    |||||
DB 1 MKKVIFVLSMFLCSQYVGOQSMHTNVEAGSINKTESIGPIDRASAASYPAHYIFHEHVAGY 60
OY 59 SESHTLYDRMTFLCLSHNTLNGACPTSENPSSSVSGETNITLQTEKRSILKRELQI 118
    |||||
DB 61 MKDHSFLDRMTFLCMSTSDSKACPTGSKSS--QGETNITLITTEKSLRKRLNLK 118
OY 119 GYKOLFKSVNCPG--PSGLTNSAHFNCNKNAASGASLYLIPAGELKNLPGGINDATL 175
    |||||
DB 119 GYKRLYESDRCHYVDKMLNSHTYKCVGSPTRGVDFLYIQGEIDGLTGIMEATL 178
OY 176 KLRVRRYSETGYTYINITIKLTDKNGIQIWLPOFKSDARVDLMLRPTGGGTYIGRNSV 235
    |||||
DB 179 ELRVKRRHYDNGHTYKNITVDLTDKNGIQIWLTPKFSRIDLNLNRPESNGKYSNVL 238
OY 236 DMCFDGYSTNSSSLEIFPODNNPKSDGKFLRKINDOTKEIAYTLLSLAGKSLTPNG 295
    |||||
DB 239 EMCLIDGYSTHSSQIETKRFDDSDQGTGNETNLTGTGEPKLPKLYKLSLGGREYFNNQ 298
OY 296 TSLNADASLETNNRITAVTAPPEISVPLVLCMPGRLQDADAKVENPEAGOVGNINVTFT 355
    |||||
DB 299 EAPFTINDSSSLFINMNRKIVSLPDISIPLVLCMPANLTFESELNPEAGYSGILNVFT 358
OY 356 PSSQTL 361
DB 359 PSSSSL 364

RESULT 4
O9XDS1 PRELIMINARY; PRT; 359 AA.
ID 09XDS1;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE TSAD protein (putative fimbrial protein).
GN TSAD OR TCED OR STY0348.

```

```

OS Salmonella typhi, and
OS Salmonella enterica.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_taxid=601, 28901;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=GIFU10007;
RA Hashimoto Y.;
RT "Identification of a putative fimbrial operon of Salmonella typhi.";
RL Submitted (JUN-1999) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=S.enterica; STRAIN=RKS 3533;
RX MEDLINE=99345391; PubMed=10417651;
RA Folkeson A., Advani A., Sukupolvi S., Pfeiffer J.D., Normark S.,
RA Lofdahl S.;
RT "Multiple insertions of fimbrial operons correlate with the evolution
of Salmonella serovars responsible for human disease.";
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher S., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Felwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrett B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar typhi CT18.";
RL Nature 413:848-852(2001).
DR EMBL: AB029403; BAA82272.1; -
DR EMBL: AJ242964; CAB51577.1; -
DR EMBL: AL627266; CAD08773.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 359 AA; 39741 MW; BB8844E9A62052CE CRC64;

Query Match 14.6%; Score 275.5; DB 16; Length 359;
Best Local Similarity 29.7%; Pred. NO. 3e-13;
Matches 90; Conservative 48; Mismatches 130; Indels 35; Gaps 11.

QY 70 FLCLSHHTLWGACPTSPNPSSSYSGEINTLTQFTKRSLLKREIQIKGYKQLFKSVN 129
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
79 WVCGRNRRNENGCACETHLLVMYAFGAASKIRLRREQSHAEITL-----ILLGSVR 131
QY 130 --CPBGLTNSAHFNCNKNAAS--GASLLYLYIPAGEELKNLFFGGIMDATLKL-RVKRRY 183
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 132 DACYAGV-----INNMAAACQNGRSLKLRIPEELAKIPTSGHWKATLVLDTLQWNG 183
QY 184 SETGTYYTINTIKITD--KGNIQMLQOF-KSDARVDLNLARPTGGGITYGNSVDMCFY 240
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 184 DDPLGTSTDTTLVNTDTHFAENAAALYFQFGATPRVDLNLARMAASQWGRANIDMCXY 243
QY 241 DGYSTNSSLEIRPDONNPKSPDGKRYLKRINDTKEIAYTSLSLAGKSLPTNGSLNI 300
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 244 DG-GYKASISLQMKIEGSKNSGSGTFQVYISDSADT--IDYASMANYGKRSIPYTRGVERSL 300
QY 301 ADAASLETNMNRITAVTMEPISVPLVLCMPGRQLQDLAK--VENPEAGQVMGINVYTFPS 357
Db 301 DNVDKAATR-----PVVLPGQRQARVCVPVPLTLTQGFENIREKKSGEQGLTVTMLMG 355
QY 358 SQT 360
Db 356 TOT 358

RESULT 5
08RGK2 PRELIMINARY; PRT; 1881 AA.
ID 08RGK2

```

AC	ORHGK2:
DT	01-JUN-2002 (Tremblrel. 21, Created)
DT	01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT	01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE	Hemolysin.
GN	FN0291.
OS	Fusobacterium nucleatum (subsp. nucleatum).
CC	Bacteria; Fusobacterium; Fusobacterium.
OX	NCBI_TaxId=76856;
RN	(1)
RP	SEQUENCE FROM N.A.
RC	STRAIN=ATCC 25586;
RX	MEDLINE=21866394; PubMed=1189109;
RA	Kapatral V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
RA	Bhatnagar V., Bartman A., Gardner W., Grecklin G., Zhu L.,
RA	Variesen O., Chu L., Kogan Y., Chaga O., Goldsman E., Bernal A.,
RA	Larsen N., D'Souza M., Malunas T., Pusch G., Haselkorn R.,
RA	Forstner M., Kyprides N., Overbeek R.;
RT	"Genome sequence and analysis of the oral bacterium Fusobacterium
RT	nucleatum strain ATCC 25586."
DR	J. Bacteriol. 184:2005-2018(2002).
DR	EMBL: AE010541; AAL94497.1; -
KW	Complete proteome.
SO	SEQUENCE 1881 AA; 204375 MW; D635155A4EFA877 CRC64;
Query Match	
Best Local Similarity 6.7%; Score 127; DB 16; Length 1881;	
Matches 86; Conservative 45; Mismatches 120; Indels 122; Gaps 19;	
QY	1 MNKILFITLFFSSLYFFFAV-----SADKIPGDESTINIFGPRDRRESSPKHN- 49
DB	5 LKKLAIEMLEPHLIIISLDGIVPDNSASKNLQVDRANGVPLVNIAP-DNNGTS--HNV 61
QY	50 -----ILNNHITAVSESHLLYDRMFLCLSHNTLNGACPTSENPSSSVSG-- 96
DB	62 YKDYVNDGREGALNN-----SKDLNLSQLGLI 89
QY	97 EFNITLQFTEKSLIKREL-----QIKGYKOLFEKSYNC---PSGLTNSAHF--NCN 144
DB	90 YGNPMLQNSKEASTIINEVSGYVKNRIEYGQITGKRVANIILANPGIYINGAGFINTGN 149
QY	145 KNAAGASGLYLITP-----AGELKNLPFGGIMPATLKLARKRKRSLEYGYTTINITIK 197
DB	150 VFTTGGRCNNMLNPKKMEIETAG--KGIDLNRINKAEELIARVAELSAPIYGGEEVVKLG 207
QY	198 LTDKNGIOIWLPOFESDARDVLDLRLPTGGGTYIGRNSYDMCFYDGYSTSSSLEIRFO-- 255
DB	208 NQGSN-----KPEYALDR-----ALGSIYAGRINT-----IYNEDGVYKTOAP 248
QY	256 -----DNNPKSDGKFFYL-----KINDTKIEIAYTLLSLAGKSLTPNTGSLNTA 301
DB	249 MYATKGVDVIVSSKGVLYLKDQAKKDKIKSSVTEIG--SKLLAENAIINISGKTSN-- 303
QY	302 DASLETETWNRIT 314
DB	304 -SGQIRAN--NNIT 314
RESULT 6	
ID	Q8VU50 PRELIMINARY; PRT; 700 AA.
AC	Q8VU50;
DT	01-MAR-2002 (Tremblrel. 20, Created)
DT	01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT	01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE	Putative polymorphic membrane protein (fragment).
OS	Chlamydia psittaci (Chlamydia phila psittaci).
OC	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX	NCBI_TaxId=83554;
RN	(1)
RP	SEQUENCE FROM N.A.
RC	STRAIN=POS;
RA	Laroucau K., Souriau A., Rodolakis A.;

RT "Isolation of a new pmp sequence and evidence of pmp polymorphism in  
RT serotype-1 Chlamydia psittaci strains."  
RT Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF243415; AAL36959.1; -  
DR InterPro: IPR003368; Chlamydia\_PMP.  
DR InterPro: IPR003357; OMP.  
DR Pfam: PF02415; DUF145; 1.  
DR Pfam: PF02385; OMP; 1.  
FT NON\_TER 700 700  
SQ SEQUENCE 700 AA; 74200 MW; 2C38E504FA8D1D1 CRC64;

Query Match 6.1%; Score 115.5; DB 2; Length 700;  
Best Local Similarity 18.3%; Pred. No. 1.3; Indels 139; Gaps 18;  
Matches 73; Conservative 63; Mismatches 124;

QY 42 NESSPKHNIINHTA-----YSESHLYDRMTFLCLIS---SHMTLNGAC----- 83  
DB 27 NEULTSDSYNGVNTSDSEFEVKETTSAGIYTCGNCVISTAGKDSPLNKSCFSETENTLS 86  
QY 84 -----PTSENPSSSSVSGETNITLQFTKRSILKRELOIKGYKQLFKSVN 129  
DB 87 FIGNGYTLCEFDNITTTASNGAIWVSGD-----QKTLNVSGFS--LFSCAH 130  
QY 130 CPSSGLT-----LNSAHF--NCKN-----NAASGASLYIYIP-----A 159  
DB 131 CPPTGTGYGAIQTKGVSTFGNNKLIIPDNKSGTGGGAIKCATGSNMLKLEGNYSYVFS 190  
QY 160 GELKNLPFGIGIMATTLARKVRRYSEYGTITIKILDKGMIQIWLPOFKSDARVDL 219  
DB 191 GNSSQKKGGAIIY--TKKLITTTADGPTLFSNNSVSAS---SPKGG-AICLDLDTSECLTA 244  
QY 220 NL-----RPTGGGTYIGRNSVDM-----CFYDGYSTNSSSLE 251  
DB 245 NLGDIPTFGDKMKVIAKTCNGSSIVKRNALIDLGSGKFTKLNAKEGPIFHPDIANTGGSTE 304  
QY 252 IRRQDNPKSD-----GR--FYLRKINDTKEIAYTL-----SLLLA 286  
DB 305 IEL--NNTESDPTTYTGKIVSGEKLSDSEKTVDPNLKSYRKQPLKIGASLYLKDGYTLE 362  
QY 287 GKSLPTNGTSLNIDAAASLETNMNRITAVTMEISVPV 325  
DB 363 AKKITQTKGSTVWMDLGTTLQTPSSSGETITLTMDINI 401

RESULT 7  
P71132 PRELIMINARY; PRT; 847 AA.  
AC P71132; 01-FEB-1997 (TREMBLrel. 02, Created)  
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)  
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
OS Chlamydia abortus.  
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia. 1.  
OX NCBI\_TaxID=83555;  
RN 11  
RP SEQUENCE FROM N.A.  
RC STRAIN=526/3;  
RX MEDLINE=96406378; PubMed=8810511;  
RA Longbottom D., Russell M., Jones G.E., Lainson F.A., Herring A.J.;  
RT "Identification of a multigene family coding for the 90 kDa proteins  
of the ovine abortion subtype of Chlamydia psittaci."  
RL FEMS Microbiol. Lett. 142:277-281(1996).  
RN 12  
RP SEQUENCE FROM N.A.  
RC STRAIN=526/3;  
RX MEDLINE=98187897; PubMed=9529048;  
RA Longbottom D., Russell M., Dunbar S.M., Jones G.E., Herring A.J.;  
RT "Molecular cloning and characterization of the genes coding for the  
highly immunogenic cluster of 90-kilodalton envelope proteins from the  
Chlamydia psittaci subtype that causes abortion in sheep."  
RL Infect. Immun. 66:1317-1324(1998).  
DR EMBL: U65942; AAC15921.1; -.

DR InterPro: IPR003368; Chlamydia\_PMP.  
DR InterPro: IPR003357; OMP.  
DR Pfam: PF02415; DUF145; 1.  
DR Pfam: PF02385; OMP; 1.  
SQ SEQUENCE 847 AA; 90694 MW; 754C958E7F1179E CRC64;

Query Match 6.1%; Score 115.5; DB 2; Length 847;  
Best Local Similarity 21.6%; Pred. No. 1.7;  
Matches 94; Conservative 58; Mismatches 153; Indels 131; Gaps 21;

QY 10 LFFSSVLTFAVSADKIPGDESTINIFGPRDRNESSPKHNIINHTAVSESHLYDRMT 69  
DB 9 LISSSLIVSNLSLSEE--PDQKTLTSAHSYNGNTNSEP-----FNPJSTSNSTGYTCG 62  
QY 70 FLCLIS---SHMTLNGACPT-----SENPSSSSVSGETNITL 102  
DB 63 NICIAVAGLDGSGLSGSCFDTAGNLSFLGNGYTLCEFDNITTTQSSHPGALSIVSG--TKTK- 120  
QY 103 QFTKRSILKRELOIKGYKQLFKSVNCPGSLTNSA-----HFNCKN 145  
DB 121 -----LDISGFS--LFSCAICPPGATGYGAIKAVGNTTIKDNLSLVFHNKCSF 166  
QY 146 NAASGASLYIYIPAGELX-----NLPR-----GIMDATIKLKVRRYSEYGTIT 192  
DB 167 GEGGAIQCKKASSSEAEKTIENQNLVFAENSSSSGGAIIYAD--KLTVSGCPTLFSNVS 225  
QY 193 NIT-----IKLIDKGMIIQIWLPOFKSDARVDLN--LRPTGGGTYIGRNSVDM----- 237  
DB 226 SASSPKGAICINDSG--ECSLTADLGDITFPDGNKIITKNGSPYTRNSIDLGSGKFT 284  
QY 238 -----CFYDGYSTNSSSLEIRFQDNPKSDOK--FYLRKINDTKEIAYTL--- 281  
DB 285 KLNAKEGFGIFFYDPTITGGSDDELINKQDPTVDT--GKIYFSEGRISDEKKAVALNKSD 343  
QY 282 -----SLLLAG-----KSLPTNGTSLNIDAAASLET--NMNRITAVTMEISV 323  
DB 344 FKQPLKIGSGLILKQGVLTLETFSFTQEGATVYMDLGTTLQTPSSSGETITLTMDIN 403  
QY 324 PVLCPMPGRLQDAKVE 339  
DB 404 ASLGGGVAPADPAKVE 419

RESULT 8  
O807H7 PRELIMINARY; PRT; 865 AA.  
AC O807H7; 01-JUN-2002 (TREMBLrel. 21, Created)  
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)  
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
OS Envelope glycoprotein (Fragment).  
GN ENV.  
OS Human immunodeficiency virus type 1.  
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN 11  
RP SEQUENCE FROM N.A.  
RC STRAIN=99CMA304;  
RX MEDLINE=21849375; PubMed=11860674;  
RA Yamaguchi J., Vallari A.S., Swanson P., Bodelle P., Kaptue L.,  
RA Nansop C., Zekeng L., Gurtler L.G., Devare S.G., Brennan C.A.;  
RT "Evaluation of HIV Type 1 Group O Isolates: Identification of Five  
Phylogenetic Clusters."  
RL AIDS Res. Hum. Retroviruses 18:269-282(2002).  
DR EMBL: AF383244; AAL98866.1; -.  
FT NON\_TER 1 1  
SQ SEQUENCE 865 AA; 97118 MW; 996B83862AA51ACA CRC64;

Query Match 6.0%; Score 114; DB 15; Length 865;  
Best Local Similarity 21.5%; Pred. No. 2.2; Indels 134; Gaps 22;  
Matches 93; Conservative 56; Mismatches 150;

QY 21 VSADKIPGDESTINIFGPRDRNESSPKHNIINHTAVSESHLYDRMTFL 71

```

Db 75 VPDPPPEYPLHN-----TDKFNITWENYVKQMGDIIDMEQSLKPCVAMTEL 125
OY 72 CUSNHT-INGACPTSENPSSSVS-----GETNITITQFTEKRLIRELOIKGYKOLER 126
Db 126 CVOONCTSLN---DTRKSNTPSPENLAKKCEFNVTIVVADKKERKQALFEVSDLMKLDK 182
OY 127 SVNCPSEGLTNSAHFNCKNNAASGA-----SLXYIPAG-----ELKNLPFG- 169
Db 183 NTNTNTMTLT-----NCNSTTISQACPKVSEFPIPIHYCAPGAYAFKCNNTIEFNGTGC 238
OY 170 -----IMDATL---KLRY-KRYSSEYV-YYTINITIKLTDK--G 202
Db 239 KNITVVTCTHGAKPTVSTQILNGLTSLKGRIRMTKNIDNVNIYTLNLTSLKTRKRG 298
OY 203 NQIWLPOKSDARDLNLPTGGGTYIGRNSVDMCFYDGYSTNSSLEIRQDNPKSD 262
Db 299 NMTVOQVSIQPMWYSMDLR-----GTERMTSRVACEY-----NSTDWE-----RTLKQT 344
OY 263 GKPYLRKINDTKELAVTSLLAG-----KSL-----TPRTGSL 298
Db 345 AERFLELVN-NTRKVDMTFSNSSGGDEPVANLHFNCHGEFFYCNTSSLNFTFSCNGSTC 403
OY 299 NTADAASLETNNRI-----TAVTMEISVPLYLCPWGRQLDPAKVENPRA 343
Db 404 NVTKSN-N-STNFRIPCRRLROYVKSWIOGSGILYAPPIRGNLTKMSNTIGMLQMDQP--- 460
OY 344 GOYMGNINVTMP 356
Db 461 ---WNHNSNATFRP 470

RESULT 9
O8XNM2 PRELIMINARY; PRT; 743 AA.
ID O8XNM2
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)
DE Hypothetical protein CPE0220.
GN CPE0220.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A.
STRAIN=13 / TYPE A;
PubMed=11792842;
RA Shima T., Ohnami K., Hirakawa H., Ohshima K., Yamashita A.,
Shima T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
"Complete genome sequence of Clostridium perfringens, an anaerobic
flesh-eater."
RT Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
RL EMBL: AP003185; BAB79926.1;
DR InterPro: IPR001899; Gram_pos_anchor.
DR TIGRfams: TIGR01167; LPTG_anchor.1.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 743 AA; 81149 MW; A7C451E9C5A5E545 CRC64;

Query Match 5.9%; Score 111.5; DB 16; Length 743;
Best Local Similarity 18.8%; Pred. No. 2.9; Mismatches 115; Indels 95; Gaps 13;
Matches 64; Conservative 66;

OY 7 IFFLFSSVLETFVAVSADRIPEDESITNIFGRDRNES---SPKHNILNNHTTYSYS 61
Db 88 LMTYVFNSSLVGEFMNIEVSAGGEALKEENKDKSITFEVSPDKVATIGLFTIMGKR 147
OY 62 HTLYDRMTFLCSSHNTLN--GACPTSENPSSSV-----SG-----ETNITIQ 103
Db 148 VELF-----LVNDMTVMTLDEAPITNAKDISVTOGDAIDLLSGVIGTDKEDSNLKY 201

```

```

OY 104 FTEKRSLIK-RELOIKGYKQLFKSVNCPSEGLTNSAHFNCKNNAASGASIVLYIPAGEL 162
Db 202 ISGDTISFVDKRAELPGVPIITYK-----YDSSGQFD----- 234
OY 163 KNLPEGIDATLKLVRKRRYSSEYTYTINITIKLTDGNIQ-----IWLPOKSDARD 218
Db 235 -----EKVYNTVYNNKKTTLGDSYTLKNTVQYVVGGMNETGSMARKYLSDSRID 285
OY 219 LNIPTGGGTYIGRNSVDMCFYDGYSTNSSLEIRQDNPKSDGKPYLRKINDTKELA 278
Db 286 IS-----NCKNTVTLTF-----NSELYAFLEKNNVTVDEKEVEAVKDKNRITK 329
OY 279 YTL-----SLLAGSLPTNGTSLINDAASLEFN 309
Db 330 FNIPDLNSDIVSTLVSMGKEVSEKTYTLNIDTAKKLEDN 369

RESULT 10
P71133 PRELIMINARY; PRT; 846 AA.
ID P71133
AC P71133;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)
DE POMP918 precursor.
OS Chlamydia abortus.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83555;
RN [1]
RP SEQUENCE FROM N.A.
STRAIN=96406378; PubMed=8810511;
RC MEDLINE=96406378; PubMed=8810511;
RA Longbottom D., Russell M., Jones G.E., Janson F.A., Herring A.J.;
RT "Identification of a multigene family coding for the 90 kDa proteins
of the ovine abortion subtype of Chlamydia psittaci."
RL FEMS Microbiol. Lett. 142:277-281(1996).
RN [2]
RP SEQUENCE FROM N.A.
STRAIN=98187897; PubMed=9529048;
RC MEDLINE=98187897; PubMed=9529048;
RA Longbottom D., Russell M., Dunbar S.M., Jones G.E., Herring A.J.;
RT "Molecular cloning and characterization of 90-kilodalton envelope proteins from the
highly immunogenic cluster of 90-kilodalton envelope proteins from the
Chlamydia psittaci subtype that causes abortion in sheep."
RL Infect. Immun. 66:1317-1324(1998).
RL EMBL: U65943; AAC15923.1;
DR InterPro: IPR003368; Chlamydia_PMP.
DR InterPro: IPR003357; OMP.
DR Pfam: PF02415; DUF145; 1.
DR Pfam: PF02385; OMP; 1.
KW Signal.
FT CHAIN 1 16 POTENTIAL.
FT SIGNAL 17 846 POMP918.
SQ SEQUENCE 846 AA; 90834 MW; 4CDC31DC0C32964E CRC64;

Query Match 5.8%; Score 110; DB 2; Length 846;
Best Local Similarity 21.9%; Pred. No. 4.4;
Matches 86; Conservative 51; Mismatches 132; Indels 124; Gaps 20;

OY 42 NESPKHNILNNHTA-----YSESHLYDMMTFCLTS---SHNTLNAC--PSENP 90
Db 27 NETLWSSDSYNGNVSDEFEVAKETTSGAITYTCEGNVCISYAGKDSFLNKSSEETENLS 86
OY 91 SSVSGET---NITLQFTEKRL---IKRELOIKGYKQLFKSVNCPSEGLTNSA--- 139
Db 87 FLG-NGYTLCEFNITQSSHGPAISVSGTNKTLDISGF-SLPSCAVCCPPTGTGAIOT 144
OY 140 -----FNCN-----KNAASGASLYIYIPAGELKLPF----- 167
Db 145 KQTTLLKDNSSLVFNKNCSTAGAGAQCKSSSTAELKL-----ENNNKLVFSENSKEKG 200
OY 168 GGIMDATLKL-----RVKRRYSSEYTYTINITIKLTDGNIQIWLPOKSDAR 216

```

Db 201 GAIYADKLTIVSGPGLFSSNNSVSHNSPKGCA-----ICIKDSGC---ECSLTANLGDIT 253  
QY 217 VDLN--LRPTGGGTYIGRNSVDM-----CFYDGYSTNSSLEIRFQDN 257  
Db 254 FDGKITTNGSGPPTVRNSIDGSGGKFKLNKKEGFIFFYPPIANTGSTIEL--N 311  
QY 258 NPKSD---GK--FYLRKINDRKELAYTL-----SLLAGKSLTP 292  
Db 312 KTESDYYTKRYFSGEKLSDEEKTVPANLKSYPEKOPKIGAGSLVLKDVLTLEAKKITQ 371  
QY 293 TNGSLNADAASLETWNMRITAVTMEISVPV 325  
Db 372 TKGSTVMDLGTTLQTPSSSGETITLNDINI 404  
RESULT 11  
Q8T2G3 PRELIMINARY; PRT; 2454 AA.  
AC 08T2G3: 01-JUN-2002 (Tremblrel. 21, Created)  
DE 01-JUN-2002 (Tremblrel. 21, Last sequence update)  
OS 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
DE Hypothetical 274.1 kDa protein.  
OS Dictyostelium discoideum (Slime mold).  
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.  
OX NCBI\_TaxID=44689;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-AX4:  
RA Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,  
Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,  
Tungann B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;  
"Sequence and Analysis of Chromosome 2 of Dictyostelium."  
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
KW EMBL: AC115608; AAL92375.1; -  
KM Hypothetical protein 2454 AA; 274123 MW; B594001B4487C724 CRC64;  
SQ SEQUENCE 2454 AA; 274123 MW; B594001B4487C724 CRC64;  
Query Match 5.8%; Score 109.5; DB 5; Length 2454;  
Best Local Similarity 23.0%; Pred. No. 19;  
Matches 79; Conservative 42; Mismatches 141; Indels 81; Gaps 18;  
QY 42 NESSPKNIILNHTTAVSESHLLYDRMTFLCLSSHNTLNGACPTSENPSSSV-SGETNI 100  
Db 996 NNSGSSGNLNNMTPTSSQTN-----SSSTTAGNPTSTSTSTGTGNTSTT 1045  
QY 101 TLOPTEKRSILKRELQKGYKQLLFKSVNCPG-----LTINS--AHFNCNNANASGA 151  
Db 1046 TTNENELPDLKK-----MEFLLLPQNGVLLPFLQSNFTTHFNVKQMLKFL 1094  
QY 152 S-LYLYIPAGELKNLPFGIWDATLKLVRKRYSEYGYTINITIKL-----TDKGNIQ 205  
Db 1095 NHVELF-----KKNPF-----TKKKLYIDSYISYIFIKLYNNYIPDQATLT 1138  
QY 206 IMLPQFSDARVDLNLRPPTGGGTYIGRN-SVDMCFYD-----GYSTNSSLEIRFQDN--- 257  
Db 1139 L-----ORLPLNIIISFS--VNKNDKISMKFYOLRTMDFLVRVNELEYEIKQKQD 1187  
QY 258 --NPKSDGKFLRKINDRKELAYTLLAGKSLPTNTSNTLNIAAASLETWNMTITA 315  
Db 1188 FLKPLSSIPFANS-SDDTSSSSSTSSL-----SLSTTTTTTTTTTKTTTTNTNSTTP 1242  
QY 316 V-TMPEISVPLCWPG-RLOLDKAVENPEAGQYGNINVTFT 356  
Db 1243 IKTGGPPTIPKLAIPGKINLGNLSGITRTDSSNNNTSSP 1285  
RESULT 12  
Q8T7V5 PRELIMINARY; PRT; 807 AA.  
AC 08T7V5: 01-JUN-2002 (Tremblrel. 21, Created)  
DE 01-JUN-2002 (Tremblrel. 21, Last sequence update)  
QY 01-JUN-2002 (Tremblrel. 21, Last sequence update)

DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
DE Nicotinic acetylcholine receptor Dalpha5 subunit.  
GN NACRAPH-34E.  
OS Drosophila melanogaster (Fruit fly).  
OC Drosophila melanogaster; Insecta;  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21969411; PubMed=11973307;  
RA Grauso M., Reenan R.A., Culetto E., Sattelle D.B.;  
RT "Novel Putative Nicotinic Acetylcholine Receptor Subunit Genes,  
RT Dalpha5, Dalpha6 and Dalpha7, in Drosophila melanogaster Identify a  
RT New and Highly Conserved Target of Adenosine Deaminase Acting on RNA-  
RT Mediated A-to-I Pre-mRNA Editing."  
RL Genetics 160:1519-1533 (2002).  
DR EMBL: AF272778; AAM13390.1; -.  
KW Receptor.  
SQ SEQUENCE 807 AA; 91223 MW; C8B4F6B34287C8C8 CRC64;  
Query Match 5.8%; Score 109; DB 5; Length 807;  
Best Local Similarity 20.2%; Pred. No. 5;  
Matches 78; Conservative 48; Mismatches 140; Indels 120; Gaps 18;  
QY 9 TLFFSVLTFEAVSADKIPDESITNIFGPRDRNESSPKNIILNH----- 54  
Db 249 TIAFISYLSGFAQLKSSSSSSSSN-----SSNNSQIILNGKHSWIFLLIYNLSAK 304  
QY 55 --ITAVSESHLLYDRMTFLCLSSHNTLNGACPTSENPSSSVSGETNITLQTEKNSLK 112  
Db 305 VCLAGYHEKRLHD-----LLDPNTLERVLNESPQLQSLFGLTLMQIIDVEKQLLV 359  
QY 113 RELQIKYKQLFKRSVCPGLTNSAHFNCKNNAAGASLYLYIPAGELKNLPFGIWD 172  
Db 360 TNWMLK-----LEWDMNLRKNTSDGVGR-----DLR-IPPRITWK 395  
QY 173 ATLKRVRKRYSEF---YGYTINITIKLTDKGNIDMLPQ--FKSDARVDLNLRPFG 226  
Db 396 PDVLM-----YNSADEFGDTGYQNVVR--NNGSC-LVYPPGFKETCKIDITWPF--- 444  
QY 227 GTYIGRNSVDMCF---YDGYSTNSSLEIRFQDNPKS-----DGKFLRKINDDTKE 276  
Db 445 --FDDQRCMKRGKSWYDGF---QLDQLQDDEGGDISYVLNGEWELLVPGKRNE 496  
QY 277 IAYT-----LSLLAGKSLPTNGTSLNTADA 303  
Db 497 IYVNCPEPIIDITFAITIRRLIYFFNLIPCVLIASMLLGLFTLPDSEKSLGVT 556  
QY 304 ASLE-TMNRITAVTMEIS--VPVL 326  
Db 557 ILLSLVFLMVAETMPATSDAVAL 582  
RESULT 13  
Q9RLA0 PRELIMINARY; PRT; 872 AA.  
AC 09RLA0: 01-MAY-2000 (Tremblrel. 13, Created)  
DE 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
DE DNA polymerase I.  
GN POLA.  
OS Rickettsia typhi.  
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
OC Rickettsiaceae; Rickettsiae; Rickettsia.  
OX NCBI\_TaxID=785;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-WTMINSTON;  
RX MEDLINE=99416441; PubMed=10486973;  
RA Anderson J.O., Anderson S.G.E.;  
RT "Genome degradation is an ongoing process in Rickettsia."

RL MOL. Biol. Evol. 16:1178-1191(1999).  
 DR EMBL: AJ238757; CAB56108.1; -  
 DR HSSP: P00582; 1KFD.  
 DR InterPro: IPR002421; 5\_3\_exonuclease.  
 DR InterPro: IPR001098; DNA\_pol.  
 DR InterPro: IPR002298; DNA\_pol.  
 DR InterPro: IPR000513; Exo\_N\_1.  
 DR InterPro: IPR003584; HNH\_2.  
 DR Pfam: PF01367; 5\_3\_exonuclease\_1.  
 DR Pfam: PF02739; 5\_3\_exonuclease\_1.  
 DR Pfam: PF00476; DNA\_pol\_A\_1.  
 DR PRINTS: PR00868; DNAPOL1.  
 DR SMART: SM00475; 53EXOC; 1.  
 DR SMART: SM00279; Hnh2; 1.  
 DR SMART: SM00482; POLAC; 1.  
 DR TIGRfams: TIGR00593; pola; 1.  
 DR PROSITE: PS00447; DNA\_POLYMERASE\_A; 1.  
 SQ SEQUENCE 872 AA; 99309 MW; BC2EBE3490879A08 CRC64;

Query Match 5.7%; Score 108; DB 2; Length 872;  
 Best Local Similarity 20.4%; Pred. No. 6; 6; Indels 112; Gaps 15;  
 Matches 77; Conservative 50; Mismatches 139;

QY 2 NRIPLFTLFFSSVLTFAVSADKIPGDESI-----TNIFGPRDRNESSPKH 48  
 DB 383 NRSIKKITYSALKHLKVFANQSHKITAIEDLMQYTLASGLVQKNLFA-----ETLTKD 437  
 QY 49 NLLNNHITVSSSHLYDMTFLCLSSHNTLN-----GACPTSSNPSSSVSGCTNIT 101  
 DB 438 NINNSAKIVINISLY-KQTLALOKNAKAFRLREIDLPCTGFIIDMKKGIKKVDANLY 496  
 QY 102 LQFTK-RSLKRELOI-----KGKQLEKFSVNCPSGLFLNSAHFNCNK 146  
 DB 497 HOLSDPEGEIILKIEEITLALSGTKFNIAQKQLEKFKKQOLSGNTLATSSITK- 555  
 QY 147 AASGASLYLYIPAGELKNLPFGGIMDATLKLK-----VKRRYSEYGYTITNITKL- 198  
 DB 556 -----AGILKLSSEGYHATILLRMRQILKKNYTSBPLQINNITRIHT 603  
 QY 199 -----TDKGNIQIWLPOKSDARVDLNRPTGGGTYIGRNSVDMCFY--DGY---STNS 247  
 DB 604 TELQSTTTLGRLSQEPNLQ-----NIPRSSDG-----NRIQAFIEDGKILISADY 652  
 QY 248 SLELEIRF-----ODNNPKSDGKFKYLRKINDTKE-----LAY 279  
 DB 653 SOIELEIILSHIANVDYLKCAFTNKEDIHTQACQIFNLQKHELTSEHRRKAKAINFGIY 712  
 DB 280 TSLSLAGKSLTPNGTS 297  
 DB 713 GISAFGLAKQLNVSNGTA 730

RESULT 14  
 ID 025579 PRELIMINARY; PRT: 2529 AA.  
 AC 025579;  
 DT 01-JAN-1998 (TREMBLrel. 05, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Toxin-like outer membrane protein.  
 GN HP0922.  
 OS Helicobacter pylori (Campylobacter pylori).  
 OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;  
 OC Helicobacter.  
 OX NCBI\_TaxID=210;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=26695 / ATCC 700392;  
 RX MEDLINE=97394467; PubMed=9252185;  
 RA \*Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,  
 RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,  
 RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,  
 RA Loftus B., Richardson D., Dodson R., Khalek H.G., Glodek A.,

RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,  
 RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,  
 RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Wathey L., Wallin E.,  
 RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,  
 RA Venter J.C.;  
 RT "The complete genome sequence of the gastric pathogen Helicobacter  
 RT pylori."  
 RL Nature 388:539-547(1997).  
 DR EMBL: AE000602; AAD07969.1; -  
 DR TIGR: HP0922;  
 DR InterPro: IPR004311; Put\_vacuole\_cyttox.  
 DR Pfam: PF03077; VacA2; 3.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 2529 AA; 274561 MW; 440882E864472EC CRC64;

Query Match 5.7%; Score 108; DB 16; Length 2529;  
 Best Local Similarity 22.5%; Pred. No. 25;  
 Matches 84; Conservative 50; Mismatches 122; Indels 118; Gaps 22;

QY 85 TSENPSSSVSGETNTTLOFTEKRSIL-----KRELOIKGYKQLFK---SVNCP 131  
 DB 977 TSNFNATYQLGNTNFTL---SSQSLNFGDPTLQNNANITLGNKSQAFAFNLSLTDNN 1033  
 QY 132 SGLTL-NSAHFNCNKNA--SGASLYLYPA-GELKNLPFGGIMDATLKLRYRRYSEY 187  
 DB 1034 SNLSLDNQSIVLANNTSARFNQASLNIYGSQATFSLFNG---GTLSSNASKLNASN 1090  
 QY 188 GYTTITITIKLTD-----KGNIQIWLPOKSDARVDLNRPTGGGTYIGRNSVD 236  
 DB 1091 ASFSNNTTINLDSVLSASNTSLNANI-----NFGASQADF---CGNTITPTASFN 1140  
 QY 237 MCFYDGYSTNSSSL---EIRQDNNPK-----SDGKFLRKIN---DDT 274  
 DB 1141 --FDSASSLNFNNLFRNGALNFRNGYPSLTKALMSVSGFVLGNNDINLSINIFDNT 1198  
 QY 275 KEIAYTSLLAGKSLTPNGTS-----LNIDAAASLETN---NRRITAVTMEI 321  
 DB 1199 KSVTY--NLLNMQKGTIGSGANGYKILFYGMKQIONATYSDNNNTQMTFNLNSSQI 1256  
 QY 322 -----SVPVL-----CWPGRQLDAKYENP-----EAGY--M 347  
 DB 1257 IOESIKNGDLTTEVLNPNPSASNTIFINAPELYNQASKNPYGSYSDNAGTYLYLT 1316  
 QY 348 GNINVTFT-SSQT 360  
 DB 1317 SNIKGLETPKGSQT 1330

RESULT 15  
 ID 08XAX2 PRELIMINARY; PRT: 304 AA.  
 AC 08XAX2;  
 DT 01-MAR-2002 (TREMBLrel. 20, Created)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Putative adhesin, similar to FlmH protein.  
 GN Z2206 OR PCS2107.  
 OS Escherichia coli O157:H7.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=83334;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O157:H7 / EDU933 / ATCC 700927;  
 RX MEDLINE=21074935; PubMed=11206551;  
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
 RA Postal G., Heckett J., Klink S., Boutin A., Shao Y., Miller L.,  
 RA Grodebeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoustis K.,  
 RA Apodaca A., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
 RA Welch R.A., Blattner F.R.;  
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."  
 Nature 409:529-533(2001).

[2]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN-0157:H7 / RIMD 0509952;  
 RX MEDLINE-21156231; Pubmed-11258796;  
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
 Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
 Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,  
 Kuhara S., Shida T., Hattori M., Shinagawa H.;  
 RT "Complete genome sequence of enterohemorrhagic Escherichia coli  
 O157:H7 and genomic comparison with a laboratory strain K-12.";  
 RL DNA Res. 8:11-22(2001).  
 DR EMBL: AE005354; AAC56266.1; -  
 DR EMBL: AP002557; BAB35530.1; -  
 DR InterPro: IPR000259; Fimbrin.  
 DR Pfam: PF00419; Fimbrin; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 304 AA; 32041 MW; EED538023D95AFD5 CRC64;

Query Match 5.7%; Score 107; DB 16; Length 304;

Best Local Similarity 21.0%; Pred. No. 2;  
 Matches 75; Conservative 51; Mismatches 125; Indels 106; Gaps 19;

```

OY 3 KILF-ITLFFSSVLTFFAVSADKIPGDEST---TNIFFPRDRNESSPKHNI---LNH 54
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8 KVLFGYLLLMAGKVFAPFSCNVD--GSSSIGAGTTSVYVND--PVIOGQNLVVDLSQH 63
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 55 ITAYSESHLYDRMTFLCLSSH-NTLNGACPTSENPSSSVSGETNITLQFTEKRLIKR 113
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 64 ISCWNDYGGWD-----TDHNLVQG-----SAFAG----- 89
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 114 ELQIKYK-QLLEKSVNCPGSLTNSAHFNCNKNASGASLYIIPAGELKNLPPGGIWD 172
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 90 --SLQSYKSLYMNWVYPPPLTNTNVLIDIGDKTPMPLPLKLYI-----TPVGAAG 140
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 173 ATLK-----LRFKRRSYETGCT-----YTINITIK-----LFDKGNIQIWLPOFK 212
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 141 VYIKAGEYIARIHMKIATFLDSGNPRNFTNIIISNSVYMPGTGCTYDSRNVTNLPDP 200
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 213 SDARVDLNLRPFGGTYIGRNSVDMCFYDGYSTNSSLEIRFODNNPKSDGKFLRKIND 272
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 201 GSAEIPL-----GYVCS-SEQKLSFYLSTGTTDSARQV-FANTAP-----D 239
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 273 DFKIATYLSLLLAGK-----SLPTNGTSLNIADASLETNWNRIYAVTMEI 321
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 240 ATKASGVGVSLMRNGKILATGENVSLGTVNKSKVPLGLSATYQGTGNKVSAGTVQSV 296
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Search completed: December 4, 2002, 17:28:40  
 Time : 40 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2002, 17:21:12 : Search time 14 Seconds

(without alignments)  
1069.497 Million cell updates/sec

Title: US-09-839-894-10

Perfect score: 1886

Sequence: 1 MNKILFIETLFSSVLTFFA.....EAGGYMGINVTFTPSQTL 361

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SWISSPROT\_40:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1558.5	82.6	360	1 CFAE_ECOLI	P25734 escherichia
2	120	6.4	2660	1 YEEJ_ECO57	O8X8V7 escherichia
3	114	6.0	650	1 SAGI_YEAST	P20840 saccharomyc
4	109	5.8	2358	1 YEEJ_ECOLI	P73347 escherichia
5	108	5.7	524	1 GAS3_YEAST	O03655 saccharomyc
6	107	5.7	304	1 YDEQ_ECOLI	P75588 escherichia
7	106	5.6	671	1 ALYS_ENTRA	P37710 enterococcu
8	99	5.2	1849	1 IG44_HAEIN	P43386 haemophilus
9	98.5	5.2	608	1 YD93_METVA	O58788 methanococc
10	98	5.2	758	1 SP21_YEAST	P35209 saccharomyc
11	98	5.2	1157	1 NI33_YEAST	P36161 saccharomyc
12	95.5	5.1	713	1 CDG2_PAEMA	P30775 penicillium
13	95	5.0	413	1 RPIW_YEAST	P30775 saccharomyc
14	95	5.0	2334	1 WAPA_BACSU	O07833 bacillus su
15	94.5	5.0	959	1 NI00_YEAST	O02629 saccharomyc
16	94	5.0	471	1 LEU2_BUCRP	P48573 buchnera ap
17	93.5	5.0	412	1 THBG_SHEEP	P50450 ovine arties
18	93	4.9	1167	1 C1AA_BACTU	P56956 bacillus th
19	92.5	4.9	320	1 CTF_CYACA	O96154 cyanidium c
20	92.5	4.9	544	1 AGM1_CANAL	O944V2 candida alb
21	92.5	4.9	1061	1 OAR_MYXXA	P36370 candida alb
22	92.5	4.9	2278	1 FABI_YEAST	P34756 saccharomyc
23	91.5	4.9	869	1 CFAE_ECOLI	P25733 escherichia
24	91.5	4.9	1783	1 Y468_MYCGE	O4460 mycoplasma
25	90.5	4.8	309	1 ELVB_CLOPE	P01556 clostridium
26	90.5	4.8	863	1 KCM4_XENLA	P30664 xenopus lae
27	90.5	4.8	1019	1 ENTK_HUMAN	P098073 homo sapien
28	90.5	4.8	1328	1 HUS2_SCHPO	O09801 schizosach
29	90.5	4.8	2193	1 POLG_HE7IM	O66479 human enter
30	89.5	4.7	1113	1 MGA2_YEAST	P40578 saccharomyc
31	89	4.7	465	1 ANT3_MOUSE	P32261 mus musculu
32	88.5	4.7	455	1 MURF_BUCAI	P57315 buchnera ap
33	88	4.7	1394	1 HAP_HAEIN	P45387 haemophilus

34	87.5	4.6	468	1 VGLC_HSYER	P12889 equine herp
35	87.5	4.6	792	1 UBPA_YEAST	P53874 saccharomyc
36	87	4.6	711	1 ERF2_SPKRA	O09892 shope fibro
37	87	4.6	1290	1 BXCL_CLOBO	P18640 clostridium
38	86.5	4.6	411	1 THBG_BOVIN	O9c136 bos taurus
39	86.5	4.6	1002	1 SPHR_AMEPV	P29815 ansacta moo
40	86.5	4.6	1273	1 WEI1_YEAST	P38968 saccharomyc
41	86.5	4.6	1567	1 RMI_DROME	O9v7b4 drosophilla
42	86.5	4.6	1769	1 YK9_YEAST	P42945 saccharomyc
43	86	4.6	368	1 YB11_YEAST	P38180 saccharomyc
44	86	4.6	493	1 FLIC_SALPA	P06178 salmoneilla
45	86	4.6	828	1 MRKC_KLEPN	P21647 klebsiella

## ALIGNMENTS

RESULT 1				
CFAE_ECOLI	STANDARD:	PRT:	360 AA.	
AC P25734:				
DT 01-MAY-1992 (Rel. 22, Created)				
DT 01-MAY-1992 (Rel. 22, Last sequence update)				
DT 01-AUG-1992 (Rel. 23, Last annotation update)				
DE CFA/I fimbrial subunit E (Colonization factor antigen I subunit E).				
GN CFAE.				
OS Escherichia coli.				
OG Plasmid NRP513.				
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;				
OC Escherichia.				
OX NCBI_TaxID=562;				
RN [1]				
RP SEQUENCE FROM N.A.				
RC STRAIN=Enterotoxigenic;				
RX MEDLINE=89330163; PubMed=2569152;				
RA Hamers A.M., Pel H.J., Willems G.A., Kusters J.G.,				
RA van der Zeijst B.A.M., Gaastra W.;				
RT "The nucleotide sequence of the first two genes of the CFA/I fimbrial				
RT operon of human enterotoxigenic Escherichia coli.,"				
RL Microb. Pathog. 6:297-309(1989).				
RM [2]				
RN SEQUENCE FROM N.A.				
RP MEDLINE=92329981; PubMed=1352712;				
RA Jordi B.J.A.M., Willems G.A., van der Zeijst B.A.M., Gaastra W.;				
RT "The complete nucleotide sequence of region I of the CFA/I fimbrial				
RT operon of human enterotoxigenic Escherichia coli.,"				
RL DNA Seq. 2:257-263(1992).				
CC -----				
CC This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC the European Bioinformatics Institute. There are no restrictions on its				
CC use by non-profit institutions as long as its content is in no way				
CC modified and this statement is not removed. Usage by and for commercial				
CC entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>				
CC or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).				
CC -----				
DR EMBL: M55661; AAC4147.1; -				
KW Antigen; Fimbria; Plasmid.				
SO SEQUENCE 360 AA; 39903 MW; 691509B63A8E69CE CRC64;				
Query Match 82.6%; Score 1558.5; DB 1; Length 360;				
Best Local Similarity 81.2%; Pred. No. 4.7e-113;				
Matches 293; Conservative 31; Mismatches 36; Indels 1; Gaps 1;				
OY 1 MNKILFIETLFSSVLTFFAASADRIKPGDESTINIFGPRDRNESSPKINILNHTAYSE 60				
DB 1 MNKILFIETLFSSVLTFFAASADRIKPGDESTINIFGPRDRNESSPKINILNHTAYSE 60				
OY 61 SHITVDRMTPLCLSSHTNLNGACPTSSSSVSGENITITQFEKSLTKREIQTIGY 120				
DB 61 SHITVDRMTPLCLSSHTNLNGACPTSSSSVSGENITITQFEKSLTKREIQTIGY 120				
OY 121 KOLFEKSVNCPGSLTILNSAHFNCNKRNAASGASLTLYIPAGELKNLPFGIMDATILKRVK 180				

```

DB 121 KQFLFNANCPKSLALNSHFFQCNREQASGATSLYIPAGELKKLPFGGVMNAVLIKINVK 180
OY 181 RRSSEYGYTYITITIKTLFGKGNIOIWLPOFKSDARYDLRLPTGGGTYGRNSVDMCFY 240
DB 181 RRDYTYGYTYITITITNLNLDKGNIOIWLPOFKSNARVDLRLPTGGGTYGRNSVDMCFY 240
OY 241 DGSTNSSSLERFODNPNKSDGKFLRKINDTKEIAYTLISLLAGKSLTPNGTSLNT 300
DB 241 DGSTNSSSLERFODNPNKSDGKFLRKINDTKEIAYTLISLLAGKSLTPNGTSLNT 300
OY 301 ADASLETMNRITAYTMEPTISVPLCWGRLOLDKAVENPEAGQINGNINVTFTSSQT 360
DB 301 -NTASLETMNRITAYTMEPTISVPLCWGRLOLDKAVENPEAGQINGNINVTFTSSQT 359
OY 361 L 361
DB 360 L 360

JLT 2
JL E0057
ID YBEL E0057 STANDARD: PRT: 2660 AA.
AC Q8X8V7: Q8X2C0: Q8X2B9:
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein yefJ.
OS 23135 OR ECS275/ECS276.
OS Escherichia coli O157:H7.
OC Bacteria: Proteobacteria: gamma subdivision: Enterobacteriaceae:
OC Escherichia.
OX NCBI_TaxID=83334:
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=0157:H7 / EDL933 / ATCC 700927;
BX MEDLINE=21074935; PubMed=11206531;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirpatrick H.A.,
RA Posfick G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potomsis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.
RL Nature 409:529-533(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=0157:H7 / RIMD 050952;
BX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shingawa H.;
RT Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.
RL DNA Res. 8:11-22(2001).
CC -1- SIMILARITY: CONTAINS 16 BIG-1 DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE INTIMIN/INVASIN FAMILY.
CC -1- CAUTION: Ref.2 sequence differs from that shown due to a
CC frameshift in position 1315.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE005423; AAC57041.1;
CC DR EMBL: AP002559; BAB36198.1; ALT_FRAME.
CC DR EMBL: AP002559; BAB36199.1; ALT_FRAME.
CC DR InterPro: IPR003344; B19_1.

```

```

DR InterPro: IPR000601; PKD domain.
KW Hypothetical protein; Repeat; Complete proteome.
FT DOMAIN 738 834
FT DOMAIN 840 929
FT DOMAIN 931 1033
FT DOMAIN 1042 1132
FT DOMAIN 1134 1236
FT DOMAIN 1245 1335
FT DOMAIN 1337 1439
FT DOMAIN 1448 1539
FT DOMAIN 1548 1652
FT DOMAIN 1653 1750
FT DOMAIN 1751 1855
FT DOMAIN 1856 1957
FT DOMAIN 1963 2056
FT DOMAIN 2065 2156
FT DOMAIN 2157 2252
FT DOMAIN 2254 2355
SQ SEQUENCE 2660 AA; 280062 MW; 01EB92A08F5C09D2 CRC64;

Query Match
Best local similarity 20.4%; Score 120; DB 1; Length 2660;
Matches 80; Conservative 66; Mismatches 156; Indels 90; Gaps 15;

OY 13 SSVLETFEAVSADKIPGDESDITNIFGPRDRNESSPKHNIINHTAY--SESHFLYDRMT 69
DB 737 SAKIATLSASNNGVLANENANVTSVNVADEGS--NPINDHTVPAVLGSATSPNNQN 793
OY 70 FLLSSHTNLINGACPTSENSSSSVSGEINTIQTETKRLI-----K 112
DB 794 ---TAKTDVNGLA-TFEDLKSSKOEDNTEVTLNCGVKQTLIVSEVDSSTPAVDLQKSK 848
OY 113 REIQIGY-KOLLFKSVNCPSGTLTNSAHFNCKNNAASGASTLYIPAGELKNLPGGIM 171
DB 849 NEVVAAGNDASATMTATVYRBAKGLNDVKTTFVNGNAAKLSQTEVNSHD-----GIA 901
OY 172 DATLKLKVRKRSEYGYTYITITIKTLFGKGNIOIWLPOFKSDARYDLRLPTGGGTYIG 231
DB 902 TATLT-----SLKRGDYVTVASVSSGQANQOVIFIGDSTAAULTLSV-PSGDIYV-- 951
OY 232 RNVVDNCFYDGSFTNSSSE---IFRODNNPKSDGKFLRKINDTKEIAYTLISLLAG 287
DB 952 -----TNTAPLMTATLDKGNPLKDKETFTSVND-----VASRFSISNG 994
OY 288 KSLTPNGTSL-----NIDASLETMNRITAYTMEPTISVPLC 327
DB 995 KGMTDSNGTATASLTGTLAGTHMITARLANSVSDIQPMTFADVADKRAVAVVLTGSAEII 1054
OY 328 WPG--RLQDLKAVENPEAGQINGNINVTFTPS 357
DB 1055 GNGVDETTLTATVAKDP-FDMVVKLSLVFERTS 1085

RESULT 3
SAG1_YEAST
ID SAG1_YEAST STANDARD: PRT: 650 AA.
AC P20840;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Alpha-agglutinin precursor (Ag-alpha-1).
GN SAG1 OR AGAL1 OR YJR004C OR J1418.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=90014768; PubMed=2677666;
RA Lipke P.N., Wojciechowski D., Kurjan J.;
RT "Ag alpha 1 is the structural gene for the Saccharomyces cerevisiae
RT alpha-agglutinin, a cell surface glycoprotein involved in cell-cell
RT interactions during mating."

```

[illegible]

Query Match	Best Local Similarity	6.0%, Score 114; DB 1; Length 650;
Matches 98; Conservative 52; Mismatches 147; Indels 172; Gaps 23;		
QY 4 ILFIPTLFSSVL-----TFPA-----VSADKIPG-----DESITNIFGRDRNE--- 43	FT CARBOHYD 329 329 O-LINKED (MAN. . .)	
DB 8 ILMLFSLALASININDITFSMLETTPILANKQPOGWTAFDFSIADASSIREGDEFTL 67	FT CARBOHYD 331 331 O-LINKED (MAN. . .)	
QY 44 SSP---KHNILNNHTA-----YSESHLYDRMTFLC-----LSSHTLN 80	FT CARBOHYD 334 334 O-LINKED (MAN. . .)	
DB 68 SMPHYRIKLNLSSQGTATISLADGTEARKCVCYQQAALVLENTFTPTCAQNDLSSTYNTID 127	FT CARBOHYD 335 335 O-LINKED (MAN. . .)	
QY 81 GACPISEMPSSSVSGENITLQTFEKSLSLRKRLQIGYKOLFKSYNC-PSGLTILNSA 139	FT CARBOHYD 338 338 O-LINKED (MAN. . .)	
DB 128 GSITPSLNFSPDGSSEYEL-----ENAKFFSGPMLVKLGQMSDYVNFDPAAFTENVF 182	FT CARBOHYD 339 339 O-LINKED (MAN. . .)	
QY 140 HFNCKNNAASGASLYL-YIPAGELKNLPFGGIMPATLKLARKRYSSELYGYTINIRK 197	FT CARBOHYD 340 340 O-LINKED (MAN. . .)	
DB 183 HSGRSTGCGFESYHLGMCPCNGYF---LGG---TEKI-----DYSSNNKVDL- 225	FT CARBOHYD 341 341 O-LINKED (MAN. . .)	
QY 198 LTKDKNIOI-----WLPQKSDARVDLRLRPTGGTYIGRN---SYDCKCFYDG--- 242	FT CARBOHYD 342 342 O-LINKED (MAN. . .)	
DB 226 -DCSSVQVYSSNDENPDMMWFPOSYNDTNADV-----TCGSLMLMTIDBEKLVDEML 275	FT CARBOHYD 343 343 O-LINKED (MAN. . .)	
QY 243 -----YSTNSSLE-IRFODNN-PRSDGKF 265	FT CARBOHYD 344 344 O-LINKED (MAN. . .)	
DB 276 WYNALQSLPANVNTIDHLEFYQTCIDTIANTTYATQESTREFIYVGGRNIGTASAKSS 335	FT CARBOHYD 345 345 O-LINKED (MAN. . .)	
QY 266 YLRKINDPTKEI---AYALS-----LLAGKSLTPTNGSLNIADA 304	FT CARBOHYD 346 346 O-LINKED (MAN. . .)	
DB 336 FISTTTTDLTISINISAYSTGISIVETGNRTTSEVISHAVTTSTRLSPATTSTLTIATQTS 395	FT CARBOHYD 349 349 O-LINKED (MAN. . .)	
QY 305 SLEETNMN-----RITAVTWPETISVPLVPCPGRL 332	FT CARBOHYD 350 350 O-LINKED (MAN. . .)	
DB 396 IYSTDSTNTTGVGDTHITTSSEVDSVETISRETAASYVAAPTSTTGWTGAM 444	FT CARBOHYD 364 364 O-LINKED (GLCNAC. . .) (POTENTIAL)	
FT CARBOHYD 402 402 O-LINKED (GLCNAC. . .) (POTENTIAL)		
FT CARBOHYD 460 460 O-LINKED (GLCNAC. . .) (POTENTIAL)		
FT CARBOHYD 485 485 O-LINKED (GLCNAC. . .) (POTENTIAL)		
FT CARBOHYD 501 501 O-LINKED (GLCNAC. . .) (POTENTIAL)		
FT CARBOHYD 614 614 N-LINKED (GLCNAC. . .) (POTENTIAL)		
FT CONFLICT 449 449 S -> P (IN REF. 1).		
FT CONFLICT 556 556 K -> E (IN REF. 1).		
FT CONFLICT 581 581 V -> L (IN REF. 1).		
SO SEQUENCE 650 AA; 70339 MW; 8BBF7A1C44C93C2B CRC64;		

```

OX NCBI_TaxID=562:
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655:
RX MEDLINE-97426617; PubMed-9278503;
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick J.A., Goeden M.A., Rose D.J.,
RA Ma B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12:
RX MEDLINE-97251358; PubMed-9097040;
RA Itoh T., Alba H., Baba T., Fujita K., Hayashi K., Inada T.,
RA Isono K., Kasai H., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T.,
RA Saito N., Sempel G., Seki Y., Sivasubram S., Tagami H.,
RA Takeda J., Takemoto K., Wada C., Yamamoto Y., Horiiuchi T.;
RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 40.1-50.0 min region on the linkage map.";
RL DNA Res. 3:379-392(1996).
CC -1- SIMILARITY: CONTAINS 13 BIG-1 DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE INTIMIN/INVASIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: A5000289; AAC75042.1; ALT_INT.
DR EMBL: D50837; BAA15800.1; -.
DR EMBL: D90836; BAA15799.1; ALT_INIT.
DR Ecogen: EG13378; Yeel.
DR InterPro: IPR003344; Big_1.
DR InterPro: IPR003535; Intimin.
DR InterPro: IPR002482; LysM.
DR InterPro: IPR00601; PKD_domain.
DR Pfam: PF02369; Big_1; 13.
DR PRINTS: PRO1369; INTIMIN.
DR SMART: SM00089; PKD; 4.
DR SMART: SM00089; PKD; 4.
KM Hypothetical protein; Repeat: Complete proteome.
RN DOMAIN 738 834
RN DOMAIN 840 931
RN DOMAIN 932 1033
RN DOMAIN 1042 1137
RN DOMAIN 1146 1237
RN DOMAIN 1246 1350
RN DOMAIN 1351 1448
RN DOMAIN 1449 1533
RN DOMAIN 1534 1655
RN DOMAIN 1661 1754
RN DOMAIN 1763 1853
RN DOMAIN 1855 1950
RN DOMAIN 1952 2053
RN DOMAIN 2053 2105
RN CONFICT 105 105
RN SEQUENCE 2358 AA; 248599 MW; 232249750BF631ED CRC64;
Query Match 5.8%; Score 109; DB 1; Length 2358;
Best Local Similarity 20.0%; Pred. No. 3.1; Indels 90; Gaps 15;
Matches 78; Conservative 65; Mismatches 157;

```

```

DB 794 ----TAKTDVNGIA-TFDLSSKSKEDNTVEVLENGVKOTLIYSFVGDSSTAQVLDQSK 848
OY 113 RELQINGYQL-LFKSVNCSGTLTNSAHFNCKNAASGASLYLIPAEELKLPBGGM 171
DB 849 NEVVADGNDSVMTATVRAKGNLMDVMTVEVNSAEAKLSOTEVNSHD-----GIA 901
OY 172 DATLKRVKRRSEFYGTNTINTIKLTDKGNIQIMLPFKSPARVDLMDPTGCGTYIG 231
DB 902 TATLT-----SLKNGDVRVTASVSSGQANQVAFIDQSTALTLISV-PSGDTTV-- 951
OY 232 RNSVDMCFYDGYSTNS---SLEIRPDNNPKSGKFFYLKINDTKETIATVLSLLAG 287
DB 952 -----TNTAPQYMTATLDKNGNPLKDEITFSVPND---VASKFSISNGG 994
OY 288 KSLPTNGTSL-----NTADAASLETMNRITVATMPEISVPVC 327
DB 995 KGMTDSNGVAIASLTGLTACTHIMARLANSVSDQPMPTVADKRAVVVLTQTSKAEII 1054
OY 328 WPG--RIOLDKVENPEAGYMGNIWTF 355
DB 1055 GNGVDETTLTATVKD-SNHPVAGITVNF 1083
Db 1055 GNGVDETTLTATVKD-SNHPVAGITVNF 1083

RESULT 5
ID GAS3_YEAST STANDARD; PRT; 524 AA.
AC 003655;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE GAS3 protein precursor.
GN GAS3 OR YMR215W OR YMR261.09.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RA Dedman K., Brown D., Bowman S., Barrell B.G., Rajandream M.A.,
RA Walsh S.V.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP IDENTIFICATION.
RX MEDLINE-20529944; PubMed-11079560;
RA Pardo M., Ward M., Bains S., Molina M., Blackstock W., Gil C.,
RA Nomdela C.;
RT "A proteomic approach for the study of Saccharomyces cerevisiae cell
wall biogenesis.";
RL Electrophoresis 21:3396-3410(2000).
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
(Potential).
CC -----
CC -1- SIMILARITY: BELONGS TO THE GAS1 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z49809; CAA89930.1; -.
DR COMPUVEAST-2DPAGE: Q03655; -.
DR SGD: S0004828; GAS3.
DR InterPro: IPR004886; GAS1.
DR Pfam: PF03198; GAS1; 1.
KW Glycoprotein; Membrane; GPI-anchor; Signal.
FT SIGNAL 1 21
FT CHAIN 22 524
FT PROPEP 7 524
FT DOMAIN 458 496
FT SER-RICH.

```





[illegible]

```

RESULT 10
SP21_YEAST
ID SP21_YEAST STANDARD: PRT: 758 AA.
AC P35209:
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE SPT21 protein
GN SPT21 OR YMR179W OR YMR010.09.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=94186069; PubMed=8138180;
RA Natsoulis G., Winston F., Boeke J.D.;
RT "The SPT10 and SPT21 genes of Saccharomyces cerevisiae.";
RL Genetics 136:93-105(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
RA Submitted (JUN-1995) to the EMBL/Genbank/DBD databases.
CC -I- FUNCTION: REQUIRED FOR NORMAL TRANSCRIPTION AT A NUMBER OF LOCI IN
CC YEAST.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb.sdb.ch).
CC -----
DR EMBL; L24436; AAA35078.1; -
DR EMBL; Z49808; CAA89912.1; -
DR PIR; S47866; S47866.
DR TRANSFAC; T04376; -
DR SGD; S0004791; SPT21.
DR DOMAIN 127 144 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 672 682 ASP/GLU-RICH (ACIDIC).
FT SEQUENCE 758 AA; 84697 MW; 7DB3FCF7EE96705 CRC64;
SQ
Query Match 5.2%; Score 98; DB 1; Length 758;
Best Local Similarity 21.9%; Pred. No. 5.1;
Matches 61; Conservative 34; Mismatches 95; Indels 88; Gaps 12.
OY 42 NESSEKHNINLHITAVSESHLTLYRMPT-----LCLSHNTLNGACPTSENPPSSSVS 95
DB 166 NISMKKGRVANNQI-----PEETLEVKRLKRTKYITLRLRSGNNTTTSRISCLOMPSSL--- 218
OY 96 GETNTTLOFTEKRSRLIKRELOIKGYKOLLFKSVNCPGSLFTLNSAHFNCKNAASGASLYL 155
DB 219 --PSATLFTPEPKSLSLFTNQIK-----NSRNAATTTTIN-----NTNSGF---- 257
OY 156 YIPAEELNLPFGGIMWDATIKLRVRRRSSEYTGTTINTITLTLDKGNQIYLPOFKSDA 215
DB 258 ---VRRROTNPV---PAPKAVRTG-----SLPIW----- 280
OY 216 RYDLMLRPTGGGTYIGRRSVDMCF-----DGYSTNSLSLEIRFDNNPKPSDGKFFYLK 269
DB 281 ---MLKPNINATGTGPRNSIAHKIYLLADRKPTANQOONHQAIAVEINTLQNDNTIQTK 336
OY 270 INDD--TKELIATYLSLLAGKSLTPNGT-----SLNI 300
DB 337 IDDSYSKRFDEFLNKRKSKTKKYSFGIATIAKPKASINI 374
RESULT 11
N133_YEAST
ID N133_YEAST STANDARD: PRT: 1157 AA.

```







DB 93 -----LTGMITLTLMTISQPEVNTITAVINTSGVNTTAVHGWPRDFKKTNAFGSF 142  
OY 232 -----RNSVDMCFYDGYSTNSSLEIRPDNNPKSDGFEYLKINDDTKEIA 278  
DB 143 TDFSNLIAAASHNKTVMYDFAPNHTNPASTDPSFAENGALYNNGTTLGKSYNDTAGIF 202  
OY 279 YTLSTLLGKSLTPT-NGSLNLIADAASLETWNRTTAVTMEISVPLWCWPGRLQDLAK 337  
DB 203 HH-----NGOTDFTSTESGIYKMLYDLADLNONNNTIDSTLKESTQLWMLGVGDIREDAV 258  
OY 338 VENPEAGQ--YMGNI-----NVTFT 355  
DB 259 KHPGQWQKSYVSISSYSSANPFT 282  
RESULT 13  
REFL\_YEAST  
ID REFL\_YEAST STANDARD; PRT; 413 AA.  
P30775:  
01-JUL-1993 (Rel. 26, Created)  
01-JUL-1993 (Rel. 26, Last sequence update)  
15-JUN-2002 (Rel. 41, Last annotation update)  
DE Peptide chain release factor 1, mitochondrial precursor (MRF-1).  
GN MRF1 OR YGL143C.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
[1]  
RN N  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93117110; PubMed=1475194;  
RA Rel H.J., Maat M.J., Rep M., Griwell L.A.;  
RT "The yeast nuclear gene MRF1 encodes a mitochondrial peptide chain  
release factor and cures several mitochondrial RNA splicing  
defects."  
RT Nucleic Acids Res. 20:6339-6346(1992).  
[2]  
RN R  
RP SEQUENCE FROM N.A.  
RX STRAIN=S288C / FY1769;  
RA Voet M., Defoor E., Verhasselt P., Riles L., Robben J., Volckaert G.;  
RT "The sequence of a nearly unclonable 22.8 kb segment on the left arm  
chromosome VII from Saccharomyces cerevisiae reveals ARO2, RPL9A,  
TIF1, MRF1 genes and six new open reading frames."  
RT Yeast 13:177-182(1997).  
[1]  
RN R  
RP -1- FUNCTION: MITOCHONDRIAL PEPTIDE CHAIN RELEASE FACTOR THAT DIRECTS  
THE TERMINATION OF TRANSLATION IN RESPONSE TO THE PEPTIDE CHAIN  
TERMINATION CODONS UAA AND UAG.  
CC -1- SUBCELLULAR LOCATION: Mitochondrial.  
CC -1- SIMILARITY: BELONGS TO THE PROKARYOTIC AND MITOCHONDRIAL RELEASE  
FACTORS FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: X60381; CAA42932.1; -;  
DR EMBL: X99960; CAA68219.1; -;  
DR EMBL: Z72665; CAA96855.1; -;  
DR PIR: S28602; S28602.  
DR SGD: S000311; MRF1.  
DR InterPro: IPR005139; PCRF.  
DR InterPro: IPR000352; pep\_rel\_factor\_I.  
DR Pfam: PF00472; RF-1; 1.  
DR Pfam: PF03462; PCRF; 1.  
DR PROSITE: PS00745; RF\_PROK\_I; 1.  
KW Protein biosynthesis; Mitochondrion; Transist peptide.  
FT TRANSIT 1 MITOCHONDRION (POTENTIAL).

FT CHAIN ? 413 PEPTIDE CHAIN RELEASE FACTOR 1.  
SQ SEQUENCE 413 AA; 46770 MW; AAAD4829748C7604 CXC64;  
Query Match 5.0%; Score 95; DB 1; Length 413;  
Best Local Similarity 21.8%; Pred. No. 4;  
Matches 67; Conservative 47; Mismatches 121; Indels 72; Gaps 14;  
OY 90 SSSVSGEN--ITLQTFKRSLI-----KRELQIKGYQLFKSVNCPSSGLTMSAHN 142  
DB 30 TSTTNKSKNSGSIPTQYTELSPLLVQAERKYEAEKLD-----LDKDLSC-----GIHPD 78  
OY 143 CNK-----NAASASLYLYIPAGELKNLPFGGIMDATLTKRVRRSE--TYGTYTI 192  
DB 79 VAKQKHAKLSALTDFTIEFKELNELKSLQEMTYSDEPSLRFADEVALVPOETISS 138  
OY 193 NITIKLTKDKNITQIWLPOKSPDARVDLNPRTGG--TYIGRNSVDMCFYDGYSTNSS 249  
DB 139 RLVNKLTP-----PHFFADKPSLLELRPGVGIEAMIFQNLLDM--YIGYA-NYRK 187  
OY 250 LEIRPDNNPKSDGFEYLKINDDTKEIAYTLSSLAKR-----SLPTNGTSLNIADA 303  
DB 188 WYRIIRISKNESSGGIIDLALISIEPAGSYDRLREFAGVHRVQRIPTSTTKG----- 239  
OY 304 ASLETWNRTTAVTMEI-----SVPLWCWPGRLQDLAKVENPEAGQYMGN----I 350  
DB 240 ---RHTSTAAVVLVPLQIDESAKSIDAYERTFKREIYVDIMRASGKGQGHVNTTDSAV 296  
OY 351 NVTFPPS 357  
DB 297 RLTHIPS 303  
RESULT 14  
WAPA\_BACSU  
ID WAPA\_BACSU STANDARD; PRT; 2334 AA.  
AC 007833;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Wall-associated protein precursor.  
GN WAPA OR N17G.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=1423;  
[1]  
RN R  
RP SEQUENCE FROM N.A.  
RX STRAIN=168;  
RC MEDLINE=93302506; PubMed=8316082;  
RA Foster S.J.;  
RT "Molecular analysis of three major wall-associated proteins of  
RT Bacillus subtilis 168: evidence for processing of the product of a  
RT gene encoding a 258 kDa precursor two-domain ligand-binding  
RT protein."  
RT Mol. Microbiol. 8:299-310(1993).  
[2]  
RN R  
RP SEQUENCE FROM N.A.  
RX STRAIN=168 / BGSC1A1;  
RC MEDLINE=97124196; PubMed=8969509;  
RA Yoshida K.-I., Sano H., Seki S., Oda M., Fujimura M., Fujita Y.;  
RA Yoshida K.-I., Shindo K., Sano H., Seki S., Fujimura M., Yanai N.,  
RA Miwa Y., Fujita Y.;  
RT "Sequencing of a 65 kb region of the Bacillus subtilis genome  
RT containing the ltc and cel loci, and creation of a 177 kb contig  
RT covering the gut-sacxy region."  
RT Microbiology 142:3113-3123(1996).  
[1]  
RN R  
RP -1- FUNCTION: STILL UNKNOWN. NOT INVOLVED IN CELL MEMBRANE METABOLISM,  
CC -----

CC MOTILITY, SECRETION OR DIFFERENTIATION.  
CC -1- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED. MAY BE RELEASED  
CC INTO THE MEDIUM.  
CC -1- DOMAIN: HAS TWO LIGAND-BINDING DOMAINS: THE N-TERMINUS, HAS THREE  
CC 101 AA REPEATS WHICH ARE RESPONSIBLE FOR CELL WALL BINDING; THE  
CC C-TERMINUS CONSISTS OF TWO BLOCKS OF RESIDUES WITH A CONSERVED  
CC MOTIF REPEATED 31 TIMES.  
CC -1- SIMILARITY: THE REPEATED MOTIF IN THE C-TERMINUS HAS SOME  
CC SIMILARITY TO THE REPEAT IN E. COLI RBS GROUP OF PROTEINS (RBSA-D).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; L05634; AAA22883.1; -  
CC EMBL; D31856; BAA06656.1; -  
CC EMBL; D29985; BAA06260.1; -  
CC EMBL; D83026; BAA11683.1; -  
CC EMBL; Z89124; CAB55959.1; -  
CC PIR; S32920; S32920.  
CC Subtilist; BG10797; wapa.  
CC InterPro; IPR003305; CBM\_Cenc.  
CC Pfam; PF02018; CBM\_4\_9; 1.  
CC Cell wall; Repeat; signal; Complete proteome.  
CC SIGNAL 1 28 OR 32 (POTENTIAL).  
CC CHAIN 29 2334 WALL-ASSOCIATED PROTEIN.  
CC DOMAIN 504 869 3 X 101 AA APPROXIMATE TANDEM REPEATS.  
CC REPEAT 504 605 1-1.  
CC REPEAT 636 736 1-2.  
CC REPEAT 769 869 1-3.  
CC DOMAIN 1021 2139 31 X 21 AA APPROXIMATE TANDEM REPEATS OF  
CC X(4)-G-X(4)-[YF]-X-D-X(2)-G-X(4).  
CC FT REPEAT 1021 1040 2-1.  
CC FT REPEAT 1042 1061 2-2.  
CC FT REPEAT 1063 1082 2-3.  
CC FT REPEAT 1083 1102 2-4.  
CC FT REPEAT 1109 1128 2-5.  
CC FT REPEAT 1129 1148 2-6.  
CC FT REPEAT 1150 1169 2-7.  
CC FT REPEAT 1174 1193 2-8.  
CC FT REPEAT 1199 1218 2-9.  
CC FT REPEAT 1219 1238 2-10.  
CC FT REPEAT 1246 1265 2-11.  
CC FT REPEAT 1267 1286 2-12.  
CC FT REPEAT 1287 1306 2-13.  
CC FT REPEAT 1307 1326 2-14.  
CC FT REPEAT 1327 1346 2-15.  
CC FT REPEAT 1347 1366 2-16.  
CC FT REPEAT 1367 1386 2-17.  
CC FT REPEAT 1387 1406 2-18.  
CC FT REPEAT 1407 1426 2-19.  
CC FT REPEAT 1427 1446 2-20.  
CC FT REPEAT 1447 1466 2-21.  
CC FT REPEAT 1467 1486 2-22.  
CC FT REPEAT 1487 1506 2-23.  
CC FT REPEAT 1507 1526 2-24 (APPROXIMATE).  
CC FT REPEAT 1527 1546 2-25.  
CC FT REPEAT 1547 1566 2-26.  
CC FT REPEAT 1567 1586 2-27.  
CC FT REPEAT 1587 1606 2-28.  
CC FT REPEAT 1607 1626 2-29.  
CC FT REPEAT 1627 1646 2-30.  
CC FT REPEAT 1647 1666 2-31.  
CC FT REPEAT 1667 1686 2-32.  
CC FT REPEAT 1687 1706 2-33.  
CC FT REPEAT 1707 1726 2-34.  
CC FT REPEAT 1727 1746 2-35.  
CC FT REPEAT 1747 1766 2-36.  
CC FT REPEAT 1767 1786 2-37.  
CC FT REPEAT 1787 1806 2-38.  
CC FT REPEAT 1807 1826 2-39.  
CC FT REPEAT 1827 1846 2-40.  
CC FT REPEAT 1847 1866 2-41.  
CC FT REPEAT 1867 1886 2-42.  
CC FT REPEAT 1887 1906 2-43.  
CC FT REPEAT 1907 1926 2-44.  
CC FT REPEAT 1927 1946 2-45.  
CC FT REPEAT 1947 1966 2-46.  
CC FT REPEAT 1967 1986 2-47.  
CC FT REPEAT 1987 2006 2-48.  
CC FT REPEAT 2007 2026 2-49.  
CC FT REPEAT 2027 2046 2-50.  
CC FT REPEAT 2047 2066 2-51.  
CC FT REPEAT 2067 2086 2-52.  
CC FT REPEAT 2087 2106 2-53.  
CC FT REPEAT 2107 2126 2-54.  
CC FT REPEAT 2127 2146 2-55.  
CC FT REPEAT 2147 2166 2-56.  
CC FT REPEAT 2167 2186 2-57.  
CC FT REPEAT 2187 2206 2-58.  
CC FT REPEAT 2207 2226 2-59.  
CC FT REPEAT 2227 2246 2-60.  
CC FT REPEAT 2247 2266 2-61.  
CC FT REPEAT 2267 2286 2-62.  
CC FT REPEAT 2287 2306 2-63.  
CC FT REPEAT 2307 2326 2-64.  
CC FT REPEAT 2327 2346 2-65.  
CC FT REPEAT 2347 2366 2-66.  
CC FT REPEAT 2367 2386 2-67.  
CC FT REPEAT 2387 2406 2-68.  
CC FT REPEAT 2407 2426 2-69.  
CC FT REPEAT 2427 2446 2-70.  
CC FT REPEAT 2447 2466 2-71.  
CC FT REPEAT 2467 2486 2-72.  
CC FT REPEAT 2487 2506 2-73.  
CC FT REPEAT 2507 2526 2-74.  
CC FT REPEAT 2527 2546 2-75.  
CC FT REPEAT 2547 2566 2-76.  
CC FT REPEAT 2567 2586 2-77.  
CC FT REPEAT 2587 2606 2-78.  
CC FT REPEAT 2607 2626 2-79.  
CC FT REPEAT 2627 2646 2-80.  
CC FT REPEAT 2647 2666 2-81.  
CC FT REPEAT 2667 2686 2-82.  
CC FT REPEAT 2687 2706 2-83.  
CC FT REPEAT 2707 2726 2-84.  
CC FT REPEAT 2727 2746 2-85.  
CC FT REPEAT 2747 2766 2-86.  
CC FT REPEAT 2767 2786 2-87.  
CC FT REPEAT 2787 2806 2-88.  
CC FT REPEAT 2807 2826 2-89.  
CC FT REPEAT 2827 2846 2-90.  
CC FT REPEAT 2847 2866 2-91.  
CC FT REPEAT 2867 2886 2-92.  
CC FT REPEAT 2887 2906 2-93.  
CC FT REPEAT 2907 2926 2-94.  
CC FT REPEAT 2927 2946 2-95.  
CC FT REPEAT 2947 2966 2-96.  
CC FT REPEAT 2967 2986 2-97.  
CC FT REPEAT 2987 3006 2-98.  
CC FT REPEAT 3007 3026 2-99.  
CC FT REPEAT 3027 3046 2-100.  
CC FT REPEAT 3047 3066 2-101.  
CC FT REPEAT 3067 3086 2-102.  
CC FT REPEAT 3087 3106 2-103.  
CC FT REPEAT 3107 3126 2-104.  
CC FT REPEAT 3127 3146 2-105.  
CC FT REPEAT 3147 3166 2-106.  
CC FT REPEAT 3167 3186 2-107.  
CC FT REPEAT 3187 3206 2-108.  
CC FT REPEAT 3207 3226 2-109.  
CC FT REPEAT 3227 3246 2-110.  
CC FT REPEAT 3247 3266 2-111.  
CC FT REPEAT 3267 3286 2-112.  
CC FT REPEAT 3287 3306 2-113.  
CC FT REPEAT 3307 3326 2-114.  
CC FT REPEAT 3327 3346 2-115.  
CC FT REPEAT 3347 3366 2-116.  
CC FT REPEAT 3367 3386 2-117.  
CC FT REPEAT 3387 3406 2-118.  
CC FT REPEAT 3407 3426 2-119.  
CC FT REPEAT 3427 3446 2-120.  
CC FT REPEAT 3447 3466 2-121.  
CC FT REPEAT 3467 3486 2-122.  
CC FT REPEAT 3487 3506 2-123.  
CC FT REPEAT 3507 3526 2-124.  
CC FT REPEAT 3527 3546 2-125.  
CC FT REPEAT 3547 3566 2-126.  
CC FT REPEAT 3567 3586 2-127.  
CC FT REPEAT 3587 3606 2-128.  
CC FT REPEAT 3607 3626 2-129.  
CC FT REPEAT 3627 3646 2-130.  
CC FT REPEAT 3647 3666 2-131.  
CC FT REPEAT 3667 3686 2-132.  
CC FT REPEAT 3687 3706 2-133.  
CC FT REPEAT 3707 3726 2-134.  
CC FT REPEAT 3727 3746 2-135.  
CC FT REPEAT 3747 3766 2-136.  
CC FT REPEAT 3767 3786 2-137.  
CC FT REPEAT 3787 3806 2-138.  
CC FT REPEAT 3807 3826 2-139.  
CC FT REPEAT 3827 3846 2-140.  
CC FT REPEAT 3847 3866 2-141.  
CC FT REPEAT 3867 3886 2-142.  
CC FT REPEAT 3887 3906 2-143.  
CC FT REPEAT 3907 3926 2-144.  
CC FT REPEAT 3927 3946 2-145.  
CC FT REPEAT 3947 3966 2-146.  
CC FT REPEAT 3967 3986 2-147.  
CC FT REPEAT 3987 4006 2-148.  
CC FT REPEAT 4007 4026 2-149.  
CC FT REPEAT 4027 4046 2-150.  
CC FT REPEAT 4047 4066 2-151.  
CC FT REPEAT 4067 4086 2-152.  
CC FT REPEAT 4087 4106 2-153.  
CC FT REPEAT 4107 4126 2-154.  
CC FT REPEAT 4127 4146 2-155.  
CC FT REPEAT 4147 4166 2-156.  
CC FT REPEAT 4167 4186 2-157.  
CC FT REPEAT 4187 4206 2-158.  
CC FT REPEAT 4207 4226 2-159.  
CC FT REPEAT 4227 4246 2-160.  
CC FT REPEAT 4247 4266 2-161.  
CC FT REPEAT 4267 4286 2-162.  
CC FT REPEAT 4287 4306 2-163.  
CC FT REPEAT 4307 4326 2-164.  
CC FT REPEAT 4327 4346 2-165.  
CC FT REPEAT 4347 4366 2-166.  
CC FT REPEAT 4367 4386 2-167.  
CC FT REPEAT 4387 4406 2-168.  
CC FT REPEAT 4407 4426 2-169.  
CC FT REPEAT 4427 4446 2-170.  
CC FT REPEAT 4447 4466 2-171.  
CC FT REPEAT 4467 4486 2-172.  
CC FT REPEAT 4487 4506 2-173.  
CC FT REPEAT 4507 4526 2-174.  
CC FT REPEAT 4527 4546 2-175.  
CC FT REPEAT 4547 4566 2-176.  
CC FT REPEAT 4567 4586 2-177.  
CC FT REPEAT 4587 4606 2-178.  
CC FT REPEAT 4607 4626 2-179.  
CC FT REPEAT 4627 4646 2-180.  
CC FT REPEAT 4647 4666 2-181.  
CC FT REPEAT 4667 4686 2-182.  
CC FT REPEAT 4687 4706 2-183.  
CC FT REPEAT 4707 4726 2-184.  
CC FT REPEAT 4727 4746 2-185.  
CC FT REPEAT 4747 4766 2-186.  
CC FT REPEAT 4767 4786 2-187.  
CC FT REPEAT 4787 4806 2-188.  
CC FT REPEAT 4807 4826 2-189.  
CC FT REPEAT 4827 4846 2-190.  
CC FT REPEAT 4847 4866 2-191.  
CC FT REPEAT 4867 4886 2-192.  
CC FT REPEAT 4887 4906 2-193.  
CC FT REPEAT 4907 4926 2-194.  
CC FT REPEAT 4927 4946 2-195.  
CC FT REPEAT 4947 4966 2-196.  
CC FT REPEAT 4967 4986 2-197.  
CC FT REPEAT 4987 5006 2-198.  
CC FT REPEAT 5007 5026 2-199.  
CC FT REPEAT 5027 5046 2-200.  
CC FT REPEAT 5047 5066 2-201.  
CC FT REPEAT 5067 5086 2-202.  
CC FT REPEAT 5087 5106 2-203.  
CC FT REPEAT 5107 5126 2-204.  
CC FT REPEAT 5127 5146 2-205.  
CC FT REPEAT 5147 5166 2-206.  
CC FT REPEAT 5167 5186 2-207.  
CC FT REPEAT 5187 5206 2-208.  
CC FT REPEAT 5207 5226 2-209.  
CC FT REPEAT 5227 5246 2-210.  
CC FT REPEAT 5247 5266 2-211.  
CC FT REPEAT 5267 5286 2-212.  
CC FT REPEAT 5287 5306 2-213.  
CC FT REPEAT 5307 5326 2-214.  
CC FT REPEAT 5327 5346 2-215.  
CC FT REPEAT 5347 5366 2-216.  
CC FT REPEAT 5367 5386 2-217.  
CC FT REPEAT 5387 5406 2-218.  
CC FT REPEAT 5407 5426 2-219.  
CC FT REPEAT 5427 5446 2-220.  
CC FT REPEAT 5447 5466 2-221.  
CC FT REPEAT 5467 5486 2-222.  
CC FT REPEAT 5487 5506 2-223.  
CC FT REPEAT 5507 5526 2-224.  
CC FT REPEAT 5527 5546 2-225.  
CC FT REPEAT 5547 5566 2-226.  
CC FT REPEAT 5567 5586 2-227.  
CC FT REPEAT 5587 5606 2-228.  
CC FT REPEAT 5607 5626 2-229.  
CC FT REPEAT 5627 5646 2-230.  
CC FT REPEAT 5647 5666 2-231.  
CC FT REPEAT 5667 5686 2-232.  
CC FT REPEAT 5687 5706 2-233.  
CC FT REPEAT 5707 5726 2-234.  
CC FT REPEAT 5727 5746 2-235.  
CC FT REPEAT 5747 5766 2-236.  
CC FT REPEAT 5767 5786 2-237.  
CC FT REPEAT 5787 5806 2-238.  
CC FT REPEAT 5807 5826 2-239.  
CC FT REPEAT 5827 5846 2-240.  
CC FT REPEAT 5847 5866 2-241.  
CC FT REPEAT 5867 5886 2-242.  
CC FT REPEAT 5887 5906 2-243.  
CC FT REPEAT 5907 5926 2-244.  
CC FT REPEAT 5927 5946 2-245.  
CC FT REPEAT 5947 5966 2-246.  
CC FT REPEAT 5967 5986 2-247.  
CC FT REPEAT 5987 6006 2-248.  
CC FT REPEAT 6007 6026 2-249.  
CC FT REPEAT 6027 6046 2-250.  
CC FT REPEAT 6047 6066 2-251.  
CC FT REPEAT 6067 6086 2-252.  
CC FT REPEAT 6087 6106 2-253.  
CC FT REPEAT 6107 6126 2-254.  
CC FT REPEAT 6127 6146 2-255.  
CC FT REPEAT 6147 6166 2-256.  
CC FT REPEAT 6167 6186 2-257.  
CC FT REPEAT 6187 6206 2-258.  
CC FT REPEAT 6207 6226 2-259.  
CC FT REPEAT 6227 6246 2-260.  
CC FT REPEAT 6247 6266 2-261.  
CC FT REPEAT 6267 6286 2-262.  
CC FT REPEAT 6287 6306 2-263.  
CC FT REPEAT 6307 6326 2-264.  
CC FT REPEAT 6327 6346 2-265.  
CC FT REPEAT 6347 6366 2-266.  
CC FT REPEAT 6367 6386 2-267.  
CC FT REPEAT 6387 6406 2-268.  
CC FT REPEAT 6407 6426 2-269.  
CC FT REPEAT 6427 6446 2-270.  
CC FT REPEAT 6447 6466 2-271.  
CC FT REPEAT 6467 6486 2-272.  
CC FT REPEAT 6487 6506 2-273.  
CC FT REPEAT 6507 6526 2-274.  
CC FT REPEAT 6527 6546 2-275.  
CC FT REPEAT 6547 6566 2-276.  
CC FT REPEAT 6567 6586 2-277.  
CC FT REPEAT 6587 6606 2-278.  
CC FT REPEAT 6607 6626 2-279.  
CC FT REPEAT 6627 6646 2-280.  
CC FT REPEAT 6647 6666 2-281.  
CC FT REPEAT 6667 6686 2-282.  
CC FT REPEAT 6687 6706 2-283.  
CC FT REPEAT 6707 6726 2-284.  
CC FT REPEAT 6727 6746 2-285.  
CC FT REPEAT 6747 6766 2-286.  
CC FT REPEAT 6767 6786 2-287.  
CC FT REPEAT 6787 6806 2-288.  
CC FT REPEAT 6807 6826 2-289.  
CC FT REPEAT 6827 6846 2-290.  
CC FT REPEAT 6847 6866 2-291.  
CC FT REPEAT 6867 6886 2-292.  
CC FT REPEAT 6887 6906 2-293.  
CC FT REPEAT 6907 6926 2-294.  
CC FT REPEAT 6927 6946 2-295.  
CC FT REPEAT 6947 6966 2-296.  
CC FT REPEAT 6967 6986 2-297.  
CC FT REPEAT 6987 7006 2-298.  
CC FT REPEAT 7007 7026 2-299.  
CC FT REPEAT 7027 7046 2-300.  
CC FT REPEAT 7047 7066 2-301.  
CC FT REPEAT 7067 7086 2-302.  
CC FT REPEAT 7087 7106 2-303.  
CC FT REPEAT 7107 7126 2-304.  
CC FT REPEAT 7127 7146 2-305.  
CC FT REPEAT 7147 7166 2-306.  
CC FT REPEAT 7167 7186 2-307.  
CC FT REPEAT 7187 7206 2-308.  
CC FT REPEAT 7207 7226 2-309.  
CC FT REPEAT 7227 7246 2-310.  
CC FT REPEAT 7247 7266 2-311.  
CC FT REPEAT 7267 7286 2-312.  
CC FT REPEAT 7287 7306 2-313.  
CC FT REPEAT 7307 7326 2-314.  
CC FT REPEAT 7327 7346 2-315.  
CC FT REPEAT 7347 7366 2-316.  
CC FT REPEAT 7367 7386 2-317.  
CC FT REPEAT 7387 7406 2-318.  
CC FT REPEAT 7407 7426 2-319.  
CC FT REPEAT 7427 7446 2-320.  
CC FT REPEAT 7447 7466 2-321.  
CC FT REPEAT 7467 7486 2-322.  
CC FT REPEAT 7487 7506 2-323.  
CC FT REPEAT 7507 7526 2-324.  
CC FT REPEAT 7527 7546 2-325.  
CC FT REPEAT 7547 7566 2-326.  
CC FT REPEAT 7567 7586 2-327.  
CC FT REPEAT 7587 7606 2-328.  
CC FT REPEAT 7607 7626 2-329.  
CC FT REPEAT 7627 7646 2-330.  
CC FT REPEAT 7647 7666 2-331.  
CC FT REPEAT 7667 7686 2-332.  
CC FT REPEAT 7687 7706 2-333.  
CC FT REPEAT 7707 7726 2-334.  
CC FT REPEAT 7727 7746 2-335.  
CC FT REPEAT 7747 7766 2-336.  
CC FT REPEAT 7767 7786 2-337.  
CC FT REPEAT 7787 7806 2-338.  
CC FT REPEAT 7807 7826 2-339.  
CC FT REPEAT 7827 7846 2-340.  
CC FT REPEAT 7847 7866 2-341.  
CC FT REPEAT 7867 7886 2-342.  
CC FT REPEAT 7887 7906 2-343.  
CC FT REPEAT 7907 7926 2-344.  
CC FT REPEAT 7927 7946 2-345.  
CC FT REPEAT 7947 7966 2-346.  
CC FT REPEAT 7967 7986 2-347.  
CC FT REPEAT 7987 8006 2-348.  
CC FT REPEAT 8007 8026 2-349.  
CC FT REPEAT 8027 8046 2-350.  
CC FT REPEAT 8047 8066 2-351.  
CC FT REPEAT 8067 8086 2-352.  
CC FT REPEAT 8087 8106 2-353.  
CC FT REPEAT 8107 8126 2-354.  
CC FT REPEAT 8127 8146 2-355.  
CC FT REPEAT 8147 8166 2-356.  
CC FT REPEAT 8167 8186 2-357.  
CC FT REPEAT 8187 8206 2-358.  
CC FT REPEAT 8207 8226 2-359.  
CC FT REPEAT 8227 8246 2-360.  
CC FT REPEAT 8247 8266 2-361.  
CC FT REPEAT 8267 8286 2-362.  
CC FT REPEAT 8287 8306 2-363.  
CC FT REPEAT 8307 8326 2-364.  
CC FT REPEAT 8327 8346 2-365.  
CC FT REPEAT 8347 8366 2-366.  
CC FT REPEAT 8367 8386 2-367.  
CC FT REPEAT 8387 8406 2-368.  
CC FT REPEAT 8407 8426 2-369.  
CC FT REPEAT 8427 8446 2-370.  
CC FT REPEAT 8447 8466 2-371.  
CC FT REPEAT 8467 8486 2-372.  
CC FT REPEAT 8487 8506 2-373.  
CC FT REPEAT 8507 8526 2-374.  
CC FT REPEAT 8527 8546 2-375.  
CC FT REPEAT 8547 8566 2-376.  
CC FT REPEAT 8567 8586 2-377.  
CC FT REPEAT 8587 8606 2-378.  
CC FT REPEAT 8607 8626 2-379.  
CC FT REPEAT 8627 8646 2-380.  
CC FT REPEAT 8647 8666 2-381.  
CC FT REPEAT 8667 8686 2-382.  
CC FT REPEAT 8687 8706 2-383.  
CC FT REPEAT 8707 8726 2-384.  
CC FT REPEAT 8727 8746 2-385.  
CC FT REPEAT 8747 8766 2-386.  
CC FT REPEAT 8767 8786 2-387.  
CC FT REPEAT 8787 8806 2-388.  
CC FT REPEAT 8807 8826 2-389.  
CC FT REPEAT 8827 8846 2-390.  
CC FT REPEAT 8847 8866 2-391.  
CC FT REPEAT 8867 8886 2-392.  
CC FT REPEAT 8887 8906 2-393.  
CC FT REPEAT 8907 8926 2-394.  
CC FT REPEAT 8927 8946 2-395.  
CC FT REPEAT 8947 8966 2-396.  
CC FT REPEAT 8967 8986 2-397.  
CC FT REPEAT 8987 9006 2-398.  
CC FT REPEAT 9007 9026 2-399.  
CC FT REPEAT 9027 9046 2-400.  
CC FT REPEAT 9047 9066 2-401.  
CC FT REPEAT 9067 9086 2-402.  
CC FT REPEAT 9087 9106 2-403.  
CC FT REPEAT 9107 9126 2-404.  
CC FT REPEAT 9127 9146 2-405.  
CC FT REPEAT 9147 9166 2-406.  
CC FT REPEAT 9167 9186 2-407.  
CC FT REPEAT 9187 9206 2-408.  
CC FT REPEAT 9207 9226 2-409.  
CC FT REPEAT 9227 9246 2-410.  
CC FT REPEAT 9247 9266 2-411.  
CC FT REPEAT 9267 9286 2-412.  
CC FT REPEAT 9287 9306 2-413.  
CC FT REPEAT 9307 9326 2-414.  
CC FT REPEAT 9327 9346 2-415.  
CC FT REPEAT 9347 9366 2-416.  
CC FT REPEAT 9367 9386 2-417.  
CC FT REPEAT 9387 9406 2-418.  
CC FT REPEAT 9407 9426 2-419.  
CC FT REPEAT 9427 9446 2-420.  
CC FT REPEAT 9447 9466 2-421.  
CC FT REPEAT 9467 9486 2-422.  
CC FT REPEAT 9487 9506 2-423.  
CC FT REPEAT 9507 9526 2-424.  
CC FT REPEAT 9527 9546 2-425.  
CC FT REPEAT 9547 9566 2-426.  
CC FT REPEAT 9567 9586 2-427.  
CC FT REPEAT 9587 9606 2-428.  
CC FT REPEAT 9607 9626 2-429.  
CC FT REPEAT 9627 9646 2-430.  
CC FT REPEAT 9647 9666 2-431.  
CC FT REPEAT 9667 9686 2-432.  
CC FT REPEAT 9687 9706 2-433.  
CC FT REPEAT 9707 9726 2-434.  
CC FT REPEAT 9727 9746 2-435.  
CC FT REPEAT 9747 9766 2-436.  
CC FT REPEAT 9767 9786 2-437.  
CC FT REPEAT 9787 9806 2-438.  
CC FT REPEAT 9807 9826 2-439.  
CC FT REPEAT 9827 9846 2-440.  
CC FT REPEAT 9847 9866 2-441.  
CC FT REPEAT 9867 9886 2-442.  
CC FT REPEAT 9887 9906 2-443.  
CC FT REPEAT 9907 9926 2-444.  
CC FT REPEAT 9927 9946 2-445.  
CC FT REPEAT 9947 9966 2-446.  
CC FT REPEAT 9967 9986 2-447.  
CC FT REPEAT 9987 10006 2-448.  
CC FT REPEAT 10007 10026 2-449.  
CC FT REPEAT 10027 10046 2-450.  
CC FT REPEAT 10047 10066 2-451.  
CC FT REPEAT 10067 10086 2-452.  
CC FT REPEAT 10087 10106 2-453.  
CC FT REPEAT 10107 10126 2-454.  
CC FT REPEAT 10127 10146 2-455.  
CC FT REPEAT 10147 10166 2-456.  
CC FT REPEAT 10167 10186 2-457.  
CC FT REPEAT 10187 10206 2-458.  
CC FT REPEAT 10207 10226 2-459.  
CC FT REPEAT 10227 10246 2-460.  
CC FT REPEAT 10247 10266 2-461.  
CC FT REPEAT 10267 10286 2-462.  
CC FT REPEAT 10287 10306 2-463.  
CC FT REPEAT 10307 10326 2-464.  
CC FT REPEAT 10327 10346 2-465.  
CC FT REPEAT 10347 10366 2-466.  
CC FT REPEAT 10367 10386 2-467.  
CC FT REPEAT 10387 10406 2-468.  
CC FT REPEAT 10407 10426 2-469.  
CC FT REPEAT 10427 10446 2-470.  
CC FT REPEAT 10447 10466 2-471.  
CC FT REPEAT 10467 10486 2-472.  
CC FT REPEAT 10487 10506 2-473.  
CC FT REPEAT 10507 10526 2-474.  
CC FT REPEAT 10527 10546 2-475.  
CC FT REPEAT 10547 10566 2-476.  
CC FT REPEAT 10567 10586 2-477.  
CC FT REPEAT 10587 10606 2-478.  
CC FT REPEAT 10607 10626 2-479.  
CC FT REPEAT 10627 10646 2-480.  
CC FT REPEAT 10647 10666 2-481.  
CC FT REPEAT 10667 10686 2-482.  
CC FT REPEAT 10687 10706 2-483.  
CC FT REPEAT 10707 10726 2-484.  
CC FT REPEAT 10727 10746 2-485.  
CC FT REPEAT 10747 10766 2-486.  
CC FT REPEAT 10767 10786 2-487.  
CC FT REPEAT 10787 10806 2-488.  
CC FT REPEAT 10807 10826 2-489.  
CC FT REPEAT 10827 10846 2-490.  
CC FT REPEAT 10847 10866 2-491.  
CC FT REPEAT 10867 10886 2-492.  
CC FT REPEAT 10887 10906 2-493.  
CC FT REPEAT 10907 10926 2-494.  
CC FT REPEAT 10927 10946 2-495.  
CC FT REPEAT 10947 10966 2-496.  
CC FT REPEAT 10967 10986 2-497.  
CC FT REPEAT 10987 11006 2-498.  
CC FT REPEAT 11007 110

DR PIR: B44402; B44402.  
DR PIR: S39173; S39173.  
DR PIR: S44518; S44518.  
DR SGD: S0001551; NUP100.  
DR InterPro: IPR004325; Nucleoporin\_FG.  
DR Pfam: PF03093; Nucleoporin\_FG; 24.  
KW Nuclear protein; Transport; Repeat.  
FT DOMAIN 33 571  
G-L-F-G  
SQ SEQUENCE 959 AA: 99988 MM: D3985F9901BBA51 CRC64;

Query Match 5.0%; Score 94.5; DB 1; Length 959;  
Best Local Similarity 19.5%; Pred. No. 13;  
Matches 80; Conservative 58; Mismatches 157; Indels 115; Gaps 17;

QY 31 STINIFGPPDRNESS-----PKHNL-----NHITAYSESHLYDRMFLCLSSHNTLNG 81  
DB 284 SMSGLEFGQSMNSSTOGVFGONNNMOINGNNNSLFGKANTFSNSASGGLFGONNOOG 343  
82 ACPTSENPSSSVSG-----ETNITLOFTEKRSLIKRELOIKGYK----- 121  
DB 344 SGLFGONSGTSSGSLFGONNOKOPNTFTQSTNGIGL-FGONNNOOOSTGLFGAKPAGT 402  
QY 122 -----QLFKSVNCPGSLTLNSAHFNCNKAASGASLYLYIPAGELKNLPFG 168  
DB 403 TGLFGONSGSTQPNSLFGTTNPTS-----NTSQGNSLFG--GATKRTMMPFG 449  
QY 169 GIWDATLKLRYRRYSEYTGTTINTITIKLTDKNIQIWLPOKFS----- 213  
DB 450 GNPRTA-----NOSGSGNSLFGTKPASTTGSIFGNNTASTVPTNGLFGNNANNSTTNT 505  
QY 214 ---DARVDLNLPT--GGGY-----IGRN-----SYDMCFYDGYSTNSSSL--- 250  
DB 506 GLFGAKPDSQSPALGGGLFGSNSNSSTIGOKPVFGCTTOWTGLFGATGNSSAVGST 565  
QY 251 -EIRFQDNMPKSDGKFEYLRKINDTKEIAYTLSSLAGKSLPTNGTSLINIAAASLETN 309  
DB 566 GKLFGONNNTLNVGTQNVPPVNNNTQN-----ALLGTTAVP---SLQQApy-----TN 610  
QY 310 WNRITAVTMP-EISVPVLCMPGRLOLDAAVENPEAGQYMGINVTFTPSS 358  
DB 611 EQLFSKISIPNSITNPVKATTSKVNADMKRNSSLTSAYRLAPKPLFAPSS 660

Search completed: December 4, 2002, 17:27:58  
Job time : 19 secs

**This Page Blank (uspto)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2002, 17:26:37 : Search time 21 Seconds  
(without alignments)  
1652.596 Million cell updates/sec

Title: US-09-839-894-10  
Perfect score: 1886  
Sequence: 1 MNKILFIFTLFFSSVLETF...EAGGYMGNINVTFTPSQTL 361

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 9613442 residues  
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR\_73:\*  
2: PIR1:\*  
3: PIR2:\*  
4: PIR3:\*  
5: PIR4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	1558.5	82.6	360	D56617	cfaf protein precu
2	949.5	50.3	363	S49539	COOD protein precu
3	930.5	49.3	364	S57937	CotD protein precu
4	275.5	14.6	359	AE0541	probable fimbrial
5	120	6.4	1335	G90975	probable factor (i
6	120	6.4	2660	E85832	probable invasiv
7	114	6.0	650	S22835	alpha-agglutinin
8	109	5.8	2383	D64962	probable membrane
9	108	5.7	524	S55097	probable membrane
10	108	5.7	2529	B64635	toxin-like outer m
11	107	5.7	304	A64904	probable fimbrial
12	107	5.7	304	C90892	probable adhesin
13	107	5.7	304	F85725	probable adhesin
14	106	5.6	671	A38109	autolysin - Entero
15	105.5	5.6	843	AC2507	hypothetical prote
16	105	5.6	598	AB1236	internalin protein
17	104.5	5.5	1441	B86807	hypothetical prote
18	103.5	5.5	691	B75622	hypothetical prote
19	103.5	5.5	1269	A90267	proteinase related
20	103.5	5.5	2399	H71879	toxin-like outer m
21	102	5.4	692	G90284	hypothetical prote
22	101	5.4	1238	AH0038	probable exported
23	100	5.3	1752	T48965	hypothetical prote
24	99.5	5.3	5627	C83339	phenylalanine-tRNA
25	99	5.2	802	AH1580	probable periplasm
26	99	5.2	856	B81389	19A-specific metal
27	98.5	5.2	608	H64473	hypothetical prote
28	98.5	5.2	608	H64473	hypothetical prote
29	98.5	5.2	4919	T31105	hypothetical prote

30	98	5.2	587	2	AC1510	internalin protein
31	98	5.2	758	2	S47866	SPR21 protein - ye
32	98	5.2	1014	2	S37405	Cytotoxic necrotiz
33	98	5.2	1157	2	S38160	NBP133 protein - y
34	97.5	5.2	1797	2	F69195	cell surface glyco
35	97	5.1	454	2	T01337	hypothetical prote
36	96.5	5.1	2364	2	I40884	cytotoxin L - Clo
37	96	5.1	992	2	T28421	probable DNA-direc
38	95.5	5.1	713	1	ALBSYR	Cyclomalodextrin
39	95.5	5.1	895	1	T02357	Mutator-like trans
40	95	5.0	413	1	S28602	transposase (12) B
41	95	5.0	561	2	A84113	cell wall-associat
42	95	5.0	2334	2	S32920	probable RTX fami
43	95	5.0	5188	2	B85547	outer membrane ass
44	94.5	5.0	835	2	E71691	replication licens
45	94.5	5.0	858	2	T47223	

#### ALIGNMENTS

RESULT 1									
D56617									
cfaf protein precursor - Escherichia coli plasmid NTP113									
C:Species: Escherichia coli									
C:Date: 05-Jan-1996 #sequence_revision 05-Jan-1996 #text_change 10-Dec-1999									
C:Accession: D56617									
R:Jordi, B.J.; Willshaw, G.A.; van der Zeijst, B.A.; Gastra, W.									
DNA Seq. 2, 257-263, 1992									
A:Title: The complete nucleotide sequence of region 1 of the CFA/I fimbrial operon of									
A:Reference number: A56617; MUID:92329981; PMID:1352712									
A:Accession: D56617									
A:Status: preliminary									
A:Molecule type: DNA									
A:Residues: 1-360 <JORD>									
A:Cross-references: GB:M5661; NID:q145507; PIDN:AA04147.1; PID:q145511									
A:Experimental source: enterotoxigenic strain, CFA/I-ST plasmid NTP113									
A>Note: sequence extracted from NCBI backbone (NCBIN:108960, NCBI:P:108972)									
A:Genetics:									
A:Gene: cfaf									
A:Genome: Plasmid									
C:Superfamily: Escherichia colonizing factor antigen cfaf									
Query Match									
Best Local Similarity 82.6%: Score 1558.5; DB 2; Length 360;									
Matches 293; Conservative 31; Mismatches 36; Indels 1; Gaps 1;									
QY	1	MNKLIFITLFFSSVLETF	FAVSADKIPGDESITNIFGPRDRNESSPKHNLNHTAYSE	60					
DB	1	MNKLIFITLFFSSGFFTF	FAVSADKIPGSEMNTNIGPHDRGSSPIYINLSYLAING	60					
QY	61	SHTLYDRMTFLCLSSHN	TLNGACPTSENPSSSSVGFEINITLOTFEKRSLIKRELQIKGY	120					
DB	61	SHHTLDRMSFLCLSSQN	TLNGACPTSSDAPGTATIDGETNITLOTFEKRSLIKRELQIKGY	120					
QY	121	KOLFESVNCPSGTLNSA	HFMCKNAASGASLYITPAGELKNLPFGGIMDATKLVK	180					
DB	121	KOLFKNANCPBKALNS	SHFQCNREQASGATLSYTPAGELKNLPFGGVMAVATKLVK	180					
QY	181	RRYSFTGYTITNITIK	TKDKGNIOIWLPOFKSADRVDLNLPPTGGTYIGHNSDMCFY	240					
DB	181	RRYDTTYGTITNITIN	VLNDKGNIOIWLPOFKSNARVDLNLPTGGTYIGHNSDMCFY	240					
QY	241	DGYSTNSSLETRFODN	NPKSDGKRYLKRINDYKEIAYTSLSLAGKSLPTNGTSLNI	300					
DB	241	DGYSTNSSLETRFODN	SKSGKRYLKRINDSKELYTSLLAGKSLPTNGTSLNI	300					
QY	301	ADAASLETNNMNTIT	AVTNPETISVPVLCWPGRLQAKYENPAGGYMGNINVTFTPSQOT	360					
DB	301	NTASLETNNMNTIT	AVTNPETISVPVLCWPGRLQAKYENPAGGYMGNINVTFTPSQOT	359					
QY	361	L	361						

Db 360 L 360

## RESULT 2

S49539

C:CoD protein precursor - Escherichia coli

C:Species: Escherichia coli

C:Date: 15-Feb-1996 #sequence\_revision 01-Mar-1996 #text\_change 10-Dec-1999

C:Accession: S49539

R:Prochlich, B.J.; Karakashian, A.; Melsen, L.R.; Wakefield, J.C.; Scott, J.R.

Mol. Microbiol. 12, 387-401, 1994

A:Title: Cooc and Cooc are required for assembly of CSI pill.

A:Reference number: S49538; M0ID:94344028; PMID:7915003

A:Accession: S49539

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-363 &lt;PRO&gt;

A:Cross-references: EMBL:X76908; NID:g488735; PIDN:CAA54230.1; PID:g488737

C:Superfamily: Escherichia colonizing factor antigen cfaE

## Query Match

Best Local Similarity 50.3%; Score 949.5; DB 2; Length 363;

Matches 197; Conservative 55; Mismatches 98; Indels 21; Gaps 11;

QY 3 KILFIPTLFSSVLTFFAVSADKIPGDE--STNIF-GPR-DRNESSPKHNLINHTAY 58

Db 2 KILFIPL---SIFSAVVSAGRYPTTVGNLTKSFQAPLDRSVQSPIYINFTNHVAGY 57

QY 59 SESHTLYDRMTFLCLSHNTLNGACPTSENPSSSVS-GETNTIOTTEKRSILKRELQI 117

Db 58 SLHSLDRIIVFLCTSSSNVNGACPTI--GTSVGYGTTITLOPTEKRSILKRNINL 114

QY 118 KGYKOLLEKSVNCPG--LTLNSAHFNCKNKA-ASGASLYLYIPAGELKLPFGIWDAT 174

Db 115 AGAKKPIWENQODEFSLMAMVNLKNSMCGAHGANGNLMLYIPAGINKLPFGIWEAT 174

QY 175 LKIRKRRSET---YGVYITITIKLTDKGNIQIWLPOFKSDARVDLNLPTGGTYI 230

Db 175 LILRLS-RYGEVSSHYGNTVITVDLTDKGNIQVWLPGFHSNPRVDLNLPTGNKYK 233

QY 231 GRNSVMCFDGYSTNSSLEIRPDNNPKSDKFLYRKINDOTKEIAYTLISLLAGKSL 290

Db 234 GNSLMDCFDGYSTNSDSVVIKFDONPPTNSEYVNLTKG-GETKLPYAVSLMGEKIF 292

QY 291 TPTNGSLNIAADAASTLNNRITTAIVTPEISVPLQMPRLDLDKAVENPEAGQYMGNI 350

Db 293 YPVNGQSFINDSSVLETNNRVTAVAMPVNVPLQMPARLLDNDVNPADAGQYSGOI 352

351 NVTFPPSQTL 361

353 YITFTPSVENTL 363

## RESULT 3

S57937

C:CoD protein precursor - Escherichia coli

C:Species: Escherichia coli

C:Date: 13-Jan-1996 #sequence\_revision 01-Mar-1996 #text\_change 10-Dec-1999

C:Accession: S57937

R:Prochlich, B.J.; Karakashian, A.; Melsen, L.R.; Scott, J.R.

submitted to the EMBL Data Library, January 1995

A:Description: The genes for CS2 pill of enterotoxigenic Escherichia coli and their inte

A:Reference number: S57934

A:Accession: S57937

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-364 &lt;PRO&gt;

A:Cross-references: EMBL:247800; NID:g897725; PIDN:CAA87763.1; PID:g897729

C:Genetics:

A:Start codon: TTG

C:Superfamily: Escherichia colonizing factor antigen cfaE

## Query Match

49.3%; Score 930.5; DB 2; Length 364;

Best Local Similarity 48.4%; Pred. No. 2e-62;

Matches 177; Conservative 64; Mismatches 118; Indels 7; Gaps 3;

QY 1 MNKILFIPTLFSSVLTFFAVSADKIPGDESTINTEGPRDRN--ESSPKNIIINHTAY 58

Db 1 MKKIVFLSMFLCSQYVGQSMHTNVEAGSTNKESIGPIDRSAASTPAYITHEHAGY 60

QY 59 SESHTLYDRMTFLCLSHNTLNGACPTSENPSSSVSGETNTIOTTEKRSILKRELQI 118

Db 61 NKDHSLEDRMTFLCWSSTADSKACPTGENSKS--QGETINIKLIFEKSLARKTLNLK 118

QY 119 GYKOLLEKSVNCPG--LTLNSAHFNCKNKAASGASLYLYIPAGELKLPFGIWDAT 175

Db 119 GYKRFLESRCIHYVDKMLNSHTVACVSPFRGVDFTLYIPOGETIDGILTCGIWEATL 178

QY 176 KLRVRRSEYGVYITITIKLTDKGNIQIWLPOFKSDARVDLNLPTGGTYMGNSV 235

Db 179 ELKVRKRYDINHGYKAVNLITVDLTDKGNIQVWLPKFSRIPRIDLNLPEKNGKYSGNVL 238

QY 236 DMCFYDGYSTNSSLEIRPDNNPKSDKFLYRKINDOTKEIAYTLISLLAGKSLPTNG 295

Db 239 EMLOYDSYTHSGSIEKRFQDDSGOTGNNEYNLTKGEPLKLLPYKLSLLGREFYPNNG 298

QY 296 TSLNIDAASTLNNRITTAIVTPEISVPLQMPRLDLDKAVENPEAGQYMGNIYFT 355

Db 299 EAFITNDSSLFTNMNRKISVLSPOISIPVLCWPMANLTFEWSLEINPEAGEYSGILNVTF 358

QY 356 PSSOTL 361

Db 359 PSSSSL 364

## RESULT 4

AE0541

probable fimbrial protein tcfD [imported] - Salmonella enterica subsp. enterica serov

C:Species: Salmonella enterica subsp. enterica serovar typh

A:Note: this species has also been called Salmonella typh

C:Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 09-Nov-2001

C:Accession: AE0541

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church

th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Parr

, S.; Moutle, S.; O'Gaora, P.

Nature 413, 848-852, 2001

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,

A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica se

A:Reference number: AB0502; PMID:11677608

A:Accession: AE0541

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-359 &lt;PAR&gt;

A:Cross-references: GB:AL513382; PIDN:CAD08773.1; PID:g16501589; GSPDB:GM00176

C:Genetics:

A:Gene: tcfD

## Query Match

Best Local Similarity 14.6%; Score 275.5; DB 2; Length 359;

Matches 90; Conservative 48; Mismatches 130; Indels 35; Gaps 11;

QY 70 FLCLSSHTLNGACPTSENPSSSVSGETNTIOTTEKRSILKRELQIKYKOLLKRSVN 129

Db 79 WVCRRNNENEGACEETHLWMTAFGAYSKIRLRFBOJSHABTTL-----ILGSVR 131

QY 130 --CPGSLTNSAHFNCKNKAAS--GASLYLYIPAGELKLPFGIWDATILK-RVKRRY 183

Db 132 DACYTG-----INMNAACQGRSLKLRIPSEELAKIPISGWKATIVLDYIQMG 183

QY 184 SETYGYTITIKLTD--KGNIQIWLPOF-KSDARVDLNLPTGGTYIGRNSVMCEY 240

Db 184 DDLGISTDITLNVGDHFAENAIYFPGGTATPRVDLNLHRMNASQMSGRANLDMCY 243

QY 241 DGSTNSSLEIRFQDNNPKSDKFLYRKINDOTKEIAYTLISLLAGKSLPTNGSLNI 300

Db 244 DG-GVAKARSLQMKIEGSKNSGTGFQYIKSDSADT--IDYAVSNWYGRSIPVTRGVEFSL 300

```

OY 301 ADAASLENNMRITAVTWPMEISVPLVCMGRGLQDAK---VENPAGGYMCNINVTFRPS 357
Db 301 DWVDKAAAR-----PVVLPGRQAAVCVPVPLLTITTPFNIREKRSGBEYOGTLVTMMG 355
OY 358 SQT 360
:|
Db 356 TQT 358

RESULT 5
G90975
probable factor [imported] - Escherichia coli (strain O157:H7, substrain RMD 0509952)
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: G90975
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
::Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genome
::Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: G90975
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1335 <HAY>
A:Cross-references: GB:BA000007; PIDN:BAB36198.1; PID:g13362243; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RMD 0509952
C:Genetics:
:Gene: ECs2775

```

[illegible]

A:Reference number: A85480; MUID:21074935; PMID:11206551  
A:Accession: E85822  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 17660 <SPO>  
A:Cross-references: GB:AAC05174; NTD:q12516151; PIDN:ANG57041.1; GSPDB:cN00145; UMGP:  
A:Experimental source: strain O157:H7, substrain EDL933  
C:Genetics:  
A:Gene: ZJ135

Query March 6.4%, Score 120; DB 2; Length 2660;  
Best Local Similarity 20.4%; Pred.No.2.1;  
Matches 80; Conservative 66; Mismatches 156; Indels 90; Gaps 15;

OY 13 SVLTFPAVSADKIPGDSITNIEFGRDRNSSPKHNILNNHITAY---SESHLYDPMT 69  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
Db 727 SAKITLTASNNGVLANEMANFYASVNVADGS--NPINDHYTFVALSGSATSPFNQN 793  
OY 70 FLCLSSHHTLNGACPTSENPPSSSVSGETNIITLOEFTKRSLI-----K 112  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
Db 794 ---TAKDVINGLA-TFDLKSSKQEDNTVEATLENGVSKOTLIVSFVGSDSTAQVDLQSK 848  
OY 113 RELQIKGY-KOLLPFSVNCPSGLTFLNSAHFNCKNKAAAGSALYLITPAGELKNLPFGIW 171  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
Db 849 NEVVADGDSDATMTATVTDANKNLNDVKYTFPNVASAALKISQTEVNSHD-----GIA 901  
OY 172 DATLTKARKRRRYSETGYTTINITIKLTDKNIOIWLPOFKSDARVDNLRLPTGGGTYG 231  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
Db 902 TATLT-----SLKNGDYTVTAVSSGSQAANGQVFIFIGDSTALTLTSY-PSGDITY-- 951  
OY 232 RNSVDMCYDGYSTNSSLSE----IRFODNNPKSDGKATYLRKINDOTEKLEINYTSLLAG 287  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
Db 952 -----TNTAPLHMHTATLDKNGNPLDKKETFEFSVPND---VASRFSISNSG 994  
OY 288 KSLPTNCTSL-----NIDDAASLETNMNRITAVTMPELSVPLC 327  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
Db 995 KGMTDSNGITALSLTGTLAGTHMITARLANSNSTOPTMTEVADRAVVVLQTSKAETI 1054  
OY 328 WPG-RIQLDAAKVENEAGQYMGYNINVTPTS 357  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
Db 1055 GNGVDETLLTATVKDP-FDNVYKKLSVVERFS 1085

RESULT 7  
S22835  
alpha-agglutinin - yeast (*Saccharomyces cerevisiae*)  
N.Alternate names: 22k glycoprotein; protein J118; protein YJR004C  
C.Species: *Saccharomyces cerevisiae*  
C.Date: 04-Dec-1992 #sequence\_revision 04-Dec-1992 #text\_change 29-Oct-1999  
C.Accession: S22835; S51229; A32822; S55192; S57019  
R.Hausser, K.; Tanner, W.  
FEBS Lett. 255, 290-294, 1989  
A.Title: Purification of the inducible alpha-agglutinin of *S. cerevisiae* and molecular  
A.Reference number: S22835; MUID:90005993; PMID:2676603  
A.Molecule type: DNA  
A.Accession: S22835  
A.Cross-references: EMBL:X16861; NID:g3352; PIDD:CAA34752.1; PID:g3353  
A.Accession: S51229  
A.Molecule type: Protein  
A.Residues: 20-24 <HA>  
R.Lipke, P.N.; Wojciechowiec, D.; Kurjan, J.  
Mol. Cell. Biol. 9, 3155-3165, 1989  
A>Title: AG-alpha-1 is the structural gene for the *Saccharomyces cerevisiae* alpha-agg  
A.Reference number: A32822; MUID:90014766; PMID:2677666  
A.Accession: A32822  
A.Molecule type: DNA  
A.Residues: 1-448, 'P', 450-555, 'E', 557-580, 'L', 582-650 <LIIP>  
A.Cross-references: GB:M28164; NID:g171041; PIDD:AAA34417.1; PID:g171044  
R.de Haan, M.; Smits, P.H.M.; Grievell, L.A.  
submitted to the EMBL Data Library, May 1995  
A.Reference number: S55193  
A.Accession: S55192





```

0y 265 ----PYLRINDPTEIAYTTLISLACKSLPTPTSLINLDAASLENNWR-ITAYMP 319
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 419 TTFNTIKNKDDT--ISATISYDKA-NSLNEIDVATATYAKSASTSOSSSKSLSTSP 475
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
0y 320 EISVYVLCMPGRLLDAKVENPEAGQYMGNT-NVTFPTPS 357
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 476 SSSSTGSSSSTG---SSASSSSSKRGVGNIVNFSQS 510
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

RESULT 10  
B64635  
toxin-like outer membrane protein HP0922 - *Helicobacter pylori* (strain 26655)  
C:Species: *Helicobacter pylori*  
C:Date: 09-Aug-1997 #sequence\_revision 09-Aug-1997 #text\_change 08-Oct-1999  
C:Accession: B64635  
R:ToMb, J.F.; White, O.; Kervatage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.  
Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khakhria, H.G.; Glodde, A.; McKenna  
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watney, L.  
Nature 388, 539-547, 1997  
P:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.  
Title: The complete genome sequence of the gastric pathogen *Helicobacter pylori*.  
P:Reference number: A64520; MUID:97394467; PMID:9252185  
A:Accession: B64635  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-2529 <OM>  
A:Cross-references: GB:AE000602; GB:AE000511; NID:q2314060; PIDN:AD07969.1; PID:q231400

Query Match	5.7%	Score 108	DB 2	Length 2529
Best Local Similarity	22.5%	Pred. No. 16		
Matches 84	Conservative 50	Mismatches 122	Indels 118	Gaps 22

```

QY 85 TSENPSSSSVGGENITITQTFEKRSLI-----KRELQIKCYKQLLP---SYMCP 131
Db 977 TSNFENATTQLLGNNITETLT---SSOSLNFNGDDTTLQNNANNTLGNKSOAAEFENSLITLLNN
QY 132 SGLTL-NSAHNCKNKNAA--SGASLYLYTPA-GELKNLPFGGIMWATKLRYKRRYSETY 187
Db 1034 SNLSLDNOSVLMNANTSAFNNQASLNTIYNGSATPNTSLFPNG---GTLSLNASSKLMA SN 1090
QY 188 GTYITNITIKLTD-----KGNIQIWLPOKFSDAVDLNLPTGGGTYIGRNSVD 236
Db 1091 ASFSNNTTINDDSVLASNTSLNANI---NFQGASQADF---GGNITIDTASFN 1140
QY 237 MCFYDGYSTNSSSL---EIRPDNNPK-----SDGKFLYLRKIN---DDT 274
Db 1141 --FDSASSLNNTNLTANGALNFNGYTPSLTRALMSVSGFYLVGNGGIDNLSDINIEFDIT 1198
QY 275 KEIATYLLSLLAGSKSLTPJNGTS-----LNTADAASLETN---WNRTAVTAMPEI 321
Db 1199 KSVYLT--NILMAQKGITISGANGIEKILFYGMKIQKQNTYSDNNNIQTWSFNLPLNSSOI 1256
QY 332 -----SVPLY-----CWPRQLQDAVEMP-----EAGQY--M 347
Db 1257 IOESIKNGDLTLEVLNPNDSASNTFIENIAPELYNOAKQNPOTGYSDYSDMQAGTYLYLT 1316
QY 348 GNINVTTP--SSQT 360
Db 1317 SNIKGLFPPKGSQT 1330

```

RESULT 11  
A64904  
probable fimbrial protein b1502 - *Escherichia coli* (strain K-12)  
C:Species: *Escherichia coli*  
C:Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 01-Mar-2002  
C:Accession: A64904  
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C.  
A.: Rose, D.J.; Mau, B.; Shao, Y.  
S:Science 277, 1453-1462, 1997  
A:Title: The complete genome sequence of *Escherichia coli* K-12.  
A:Reference number: A64720; MUID:97426617; PMID:9278503  
A:Accession: A64904

A:Status: nuncic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-304 <BLAT>  
A:Cross-references: GI:AE000247; GH:U00096; PIDN:AACT4576.1; PID:g1787777  
A:Experimental source: Strain K-12, substrain M61655  
C:Superfamily: fimbrial protein fliH  
C:Keywords: fimbria

Query Match	5.7%	Score 107	DB 2	Length 304
Best Local Similarity	21.0%	Pred. No. 1.2		
Matches 75	Conservative 51	Mismatches 125	Indels 106	Gaps 19

```

OY 3 KLFLEIFLEFFSVLETFVAVSADKIPGDESI-----TNIJEGPRDRNESSPKHNI---LNNH 54
Db 8 KYLEFGIYLILMAGKFAKFAISCND---GSSSICAGTTSYVNLID-PYIQPGONLVDYDSOH 63
OY 55 ITAYSESHTLXDRMTFLCLSSH-NTLNGACAPTSENPSSSVSGETNITLQFTEKSLIKR 113
Db 64 ISCNMDYGGWYD-----TDHINLVQG-----SAFAG----- 89
OY 114 ELQJQGYK-QLLFKSVNCPGGLTNSAHFNCKNNAASGLYLYIPAGELKNLPFGITMD 172
Db 90 --SDQSYRGLSLWNVNYTPPLTTNTNVLIDGDKTPMPLPLXYI-----TPGAANG 140
OY 173 ATLK-----LRYKRYSETYGF-----YTINITIK-----LTDKGNILQILPOFK 212
Db 141 VIKRGEVIARIHMVKIATLISGPNRFTWNIISNNNVYMPGCGTVDNRVATYDLPPDF 200
OY 213 SDARYDLNLPRPGGTYIGRNSVDMCEFYDGYSTNSSSLIEIRQDNNPKSDGKFYLKRTND 272
Db 201 GSAELPL-----GVYCS-SEOKLSFYLSGATYDSSHOV-FANAP-----D 239

```

RESULT 12  
C90892  
probable adhesin [imported] - Escherichia coli (strain O157:H7, substrain RIMD 0509955)  
C:Species: Escherichia coli  
C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 03-Aug-2001  
C:Accession: C90892  
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.  
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and g  
A:Reference number: A99629; MUID:21156231; PMID:11258796  
A:Accession: C90892  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-504 <HAP>  
A:Cross-references: GB:BA000007; PIDN:BAR5530.1; PID:G13361573; GSPDB:GN00154  
A:Experimental source: strain O157:H7, substrain RIMD 0509955  
C:Genetics:  
A:Gene: ECS2107  
C:Superfamily: fimbrial protein fliH

Query Match	Similarity	5.7%	Score 107	DB 2	Length 304
Best Local	Similarity	21.0%	Pred. No. 1.2		
Matches	75	Conservative	51	Mismatches	125
				Indels	106
				Gaps	19
Oy	3	KILF-IFTLFFSVLETFYAVSADKIPGDESI-----TNIFGPRDRNESSPKHNI--LNNH	54		
Db	8	KVLFEIYILLMAGKAFKFAISCND---GGSSIGAGCTTYSYVMD--PVIOPGQLVVDLSQH	63		
Oy	55	ITAVSESHLTLYDRMTFLCLSSH--NTLNGACPTSENPSSSSVSGETNITLQFTEKSLIKR	113		
Db	64	ISCNNDYGGWMD-----TDHINLVQG-----SAFAG-----	89		
Oy	114	ELQIKGYR-QLLFEKSYNCPGSLTLNSAHFNCKNKNAASASLYLYIPAGELKNLPGGIWD	172		
Db	90	--SLOSYSGLYNNVNTVPEPLTNTNVLIDGDKTPMELPKLYI-----TPGGAAGG	140		

```

OY      173 ATLK-----LRVRRRSSEYGT-----YTINILTK-----LFDGNGIOIWLPOKR 212
           :|          :|          :|          :|          :|          :|
Db      141 VVIKAGEVIAIRIMHYKATATAGSGNPRNFTHNIISNSVYMPLPGCGCTVDNRVTANLPDP 200
           :|          :|          :|          :|          :|          :|
OY      213 SPARDVLNLRPTGGGYIGRNSVDMCFDGYSTNSSSLEIRFODNNPKSDGRFYELRKIND 272
           :|          :|          :|          :|          :|          :|
Db      201 GSAEIPV-----GVYCS-SEKLSFYLSTGTTDSARQV-FANTAP-----D 239
           :|          :|          :|          :|          :|          :|
OY      273 DKPEIVTSLLLAGK-----SLPTNGTSLNIAAASLETMMNRITAVTMPEI 321
           :|          :|          :|          :|          :|          :|
Db      240 ATKASGVGSMLMKNKITLATGENVSLGTVNKSVPVGLSATYGQTGNKVSAGTVOSV 296
           :|          :|          :|          :|          :|          :|

RESULT 13
P85725
probable adhesin, fimb type protein Z2206 [imported] - Escherichia coli (strain O157:H7,
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
Accession: F85725
Server, N.T.: Plumett II, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Diallantia, E.; Petamoustis, K.; Apodaca,
Nature 409, 529-533, 2001
A>Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: AB5480; MDOI:21074935; PMID:11206551
A:Accession: F85725
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-304 <STO>
A:Cross-references: GB:AEO05174; NID:g12515169; PIDD:AAG56266.1; GSPDB:GN00145; UWGP:Z22
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z2206
C:Superfamily: fimbrial protein fimH
```

[illegible]

```

A:Reference number: A38109; MUUD:91358349; PMID:1679432
A:Accession: A38109
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-671 <BEU>
A:Cross-references: GB:M58002; NID:g153658; PIDN:AAA67325.1; PID:g829194

Query Match          5.6%; Score 106; DB 2; Length 671;
Best Local Similarity 23.6%; Pred. No. 3.9;
Matches 73; Conservative 39; Mismatches 103; Indels 94; Gaps 17;

QY 41 RNESPSPKHNIT-NNHNTTAYSESHLTLYDRMTFCLSHNTLNCACPTSENPSSSV----- 94
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 309 KATADPSTYAKANNVITAY--NLTYDTIPSSGNGTGGGTVPNGTGSNNQSGNTYTYVK 366
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 95 SGET--NTTLOF-----TEKRSL--IKRELQIKGYKQLLEKFSVNCPSGLTLNSAHENCNKN 146
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 367 SCDTLNKIAAQGVSVANLRSMWNGISGDILFVGOKLIYKKA---SGNTGGSGNCGSNNN 423
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 147 AASGASLVIYIPAGELKNLPFGGINDATLKLKRRKRYSEYTYGTYINTTIKLTQKGNQI 206
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 424 -OSGNTYTYTVKSGDTLN-----KIAAQY---VTV-----ANLRS 455
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 207 WLPEKSPDARVGLN-----RPTGGGTLYIGRNSVDMCFYDGYSTSSSLIEIFODNN 258
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 456 W-----NGISGLLIVGOKLIYKKTSGNT-----GGSSNGS-----NNN 491
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 259 PKSDKFFLYLRKINDDTKEIA--YTLSS-----LLLAGKSLPTNGTSLINTADA 303
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 492 QSGCTMYTYTIKSGDTLNKIAAQGVSVANLRSMWNGISGDILFVAGQKLIYKKTSGNTGGS 551
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 304 ASLETTNNMR 312
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 552 SNGGSGNNNO 560
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

[illegible]



**This Page Blank (uspto)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 6, 2002, 18:15:00 ; Search time 12 Seconds  
(without alignments)  
488.623 Million cell updates/sec

Title: US-09-839-894-10

Perfect score: 1886

Sequence: 1 MNKILFIFLFFSSVLFTEFA.....EAGGYMGNIYVTPSSQTL 361

Scoring table:

Gapop 10.0 , Gapext 0.5

Searched: 103943 seqs, 16242309 residues

tal number of hits satisfying chosen parameters: 103943

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /cgnt2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*  
2: /cgnt2\_6/ptodata/1/pubpaa/PTCT\_NEW\_PUB.pep:\*  
3: /cgnt2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgnt2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgnt2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgnt2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
7: /cgnt2\_6/ptodata/1/pubpaa/PTCT\_PUBCOMB.pep:\*  
8: /cgnt2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgnt2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*  
10: /cgnt2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
11: /cgnt2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:\*  
12: /cgnt2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep:\*  
13: /cgnt2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*  
14: /cgnt2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1886	100.0	361	9	US-09-839-894-10
2	1862.5	98.8	361	9	US-09-839-894-28
3	1542	81.8	359	9	US-09-839-894-29
4	937	49.7	362	9	US-09-839-894-31
5	916.5	48.6	364	9	US-09-839-894-30
6	247.5	13.1	353	9	US-09-839-894-32
7	119.5	6.3	770	10	US-09-303-232-2
8	116.5	5.2	1600	9	US-10-092-880-10
9	109	5.8	2383	10	US-09-912-020-302
10	101	5.4	1599	9	US-10-092-880-9
11	95	5.0	867	9	US-09-839-894-6
12	94.5	5.0	2353	10	US-09-797-862-33
13	91.5	4.9	1143	10	US-09-924-154-14
14	90.5	4.8	564	9	US-09-944-160-12
15	89.5	4.7	811	9	US-10-011-588-29
16	89	4.7	465	9	US-09-987-021-6
17	89	4.7	465	10	US-09-957-485-6
18	88	4.7	921	9	US-09-117-447-6
19	86.5	4.6	501	10	US-09-303-232-6

20	86.5	4.6	871	10	US-09-886-468-21	Sequence 21, Appl
21	86.5	4.6	1536	9	US-10-092-880-2	Sequence 2, Appl
22	86	4.6	3503	9	US-10-108-605-237	Sequence 237, App
23	85.5	4.5	837	10	US-09-815-242-5883	Sequence 5883, Ap
24	85.5	4.5	875	10	US-09-815-242-13080	Sequence 13080, A
25	85.5	4.5	2434	10	US-09-815-242-5835	Sequence 5835, Ap
26	85.5	4.5	6281	10	US-09-815-242-12996	Sequence 12996, A
27	84.5	4.5	543	10	US-09-871-212-7	Sequence 7, Appl
28	84.5	4.5	585	9	US-10-024-632-11	Sequence 11, Appl
29	84.5	4.5	791	10	US-09-764-870-337	Sequence 337, App
30	84.5	4.5	916	9	US-09-978-295A-390	Sequence 390, App
31	84.5	4.5	916	9	US-09-978-697-390	Sequence 390, App
32	84.5	4.5	916	12	US-10-052-586-78	Sequence 78, Appl
33	84.5	4.5	916	12	US-10-052-586-78	Sequence 1145, A
34	84.5	4.5	1032	10	US-09-741-669-304	Sequence 304, App
35	84	4.5	1325	10	US-09-815-242-11145	Sequence 11145, A
36	84	4.5	1477	9	US-10-092-880-4	Sequence 4, Appl
37	83.5	4.4	922	10	US-09-886-468-19	Sequence 19, Appl
38	83.5	4.4	5795	10	US-09-815-242-12610	Sequence 12610, A
39	83	4.4	518	9	US-09-976-297-4	Sequence 4, Appl
40	83	4.4	1349	10	US-09-815-242-5898	Sequence 5898, Ap
41	83	4.4	1349	10	US-09-815-242-13137	Sequence 13137, A
42	82.5	4.4	724	10	US-09-925-300-1053	Sequence 1053, Ap
43	82	4.3	3256	10	US-09-919-172-98	Sequence 98, Appl
44	81.5	4.3	698	10	US-09-899-471-5	Sequence 5, Appl
45	81	4.3	252	9	US-09-887-853-4	Sequence 4, Appl

#### ALIGNMENTS

RESULT 1

US-09-839-894-10  
; Sequence 10, Application US/09839894  
; Patent No. US20020176868A1

; GENERAL INFORMATION:

; APPLICANT: Altboum, Zeev

; APPLICANT: Barry, Eileen M.

; APPLICANT: Levine, Myron M.

; TITLE OF INVENTION: University of Maryland

; TITLE OF INVENTION: ISOLATION AND CHARACTERIZATION OF THE

; FILE REFERENCE: USFMD 0064

; CURRENT APPLICATION NUMBER: US/09/839,894

; CURRENT FILING DATE: 2001-04-20

; PRIOR APPLICATION NUMBER: 60/198,626

; PRIOR FILING DATE: 2000-04-20

; NUMBER OF SEQ ID NOS: 40

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 10

; LENGTH: 361

; TYPE: PRT

; ORGANISM: E. coli

US-09-839-894-10

Query Match	100.0%	Score 1886	DB 9	Length 361
Best Local Similarity	100.0%	Pred. No. 2.2e-163		
Matches 361	Conservative	0	Mismatches 0	Indels 0
Gaps 0				
QY	1	MNKLIFLFFSSVLFTEFAVSADRIKPGDESITNIFGPRDRNESSPKHNLNNHTAYSE	60	
DB	1	MNKLIFLFFLFFSSVLFTEFAVSADRIKPGDESITNIFGPRDRNESSPKHNLNNHTAYSE	60	
QY	61	SHTLDRMTFLCLSSHTNLNGACPTSENSSVSGETNITLQTEKRSLLKRELDQIGY	120	
DB	61	SHTLDRMTFLCLSSHTNLNGACPTSENSSVSGETNITLQTEKRSLLKRELDQIGY	120	
QY	121	KOLLKSVNCPGSLTNSAHFNCKNKAASGASLYITRAGEIKNLPFGIMDATIKLVK	180	
DB	121	KOLLKSVNCPGSLTNSAHFNCKNKAASGASLYITRAGEIKNLPFGIMDATIKLVK	180	
QY	181	RRYSEYTGTYTNTIKITDKGNIQIWLPOFKSDARVDLNLKPTGCGTYIGRNSDMCFY	240	
DB	181	RRYSEYTGTYTNTIKITDKGNIQIWLPOFKSDARVDLNLKPTGCGTYIGRNSDMCFY	240	

Db	181	RRYSFYGTYTINTIKTLTKDKNIGIQLWLPQKSDARVDLMLRPTGGGTYIGRNSVDMCFY	240
Qy	241	DGYSTNSSLEIRFQDNPNPKSDGKFYLRKINDDTKEIAYTLLSLACKSLTPITNGTSLNI	300
Db	241	DGYSTNSSLEIRFQDNPNPKSDGKFYLRKINDDTKEIAYTLLSLACKSLTPITNGTSLNI	300
Qy	301	ADASLTNNNRITAYVMPETISYVPLQCPGRLODAKVENDEAGQVNGNINVTPTSSQT	360
Db	301	ADASLTNNNRITAYVMPETISYVPLQCPGRLODAKVENDEAGQVNGNINVTPTSSQT	360
Qy	361	L 361	
Db	361	L 361	

```

RESULT 2
US-09-839-894-28
: Sequence 28, Application US/09839894
: Patent No. US2002017668A1
GENERAL INFORMATION:
APPLICANT: Alboum, Zeev
APPLICANT: Barry, Eliezer M.
APPLICANT: Levine, Myron M.
TITLE OF INVENTION: Isolation and Characterization of the
File of Invention: CSA OPERON
FILE REFERENCE: USFMD 006A
CURRENT APPLICATION NUMBER: US/09/839,894
CURRENT FILING DATE: 2001-04-20
PRIOR APPLICATION NUMBER: 60/198,626
PRIOR FILING DATE: 2000-04-20
NUMBER OF SEQ ID NOS: 40
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 28
LENGTH: 361
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: ETEC Protein Homology Sequence
US-09-839-894-28

```

Query Match	Similarity	Score	ID	Length
Best Local	99.4%	Pred. No. 3e-161		
Matches	359	Conservative	0	Mismatches 1; Indels 1; Gaps 1
QY	1	MNKLIFLTFFSSVLTFFFAVSADRI	IPGDESTITNIFGPRDRNESSPKHILNNHTAYSE	60
Db	2	MNKLIFLTFFSSVLTFFFAVSADRI	IPGDESTITNIFGPRDRNESSPKHILNNHTAYSE	61
QY	61	SHTLYDRMTFFCLSSHTNLNGACPF	SENNSSSVSGETITITQFTFKRLKRELOIKGY	120
Db	62	SHTLYDRMTFFCLSSHTNLNGACPF	SENNSSSVSGETITITQFTFKRLKRELOIKGY	121
QY	121	KOLLEFSVNCPSGGLTNSAHFNCN	NAASGASLYLYIPAGELKNLPFGGIMPATLKLARK	180
Db	122	KOLLEFSVNCPSGGLTNSAHFNCN	NAASGASLYLYIPAGELKNLPFGGIMPATLKLARK	181
QY	181	RRYSETTYCTYTINITIKLTDKGN	IOIWLPEQKSDARVDLNLPRTGCGTYIGRNSVDMCFY	240
Db	182	RRYSETTYCTYTINITIKLTDKGN	IOIWLPEQKSDARVDLNLPRTGCGTYIGRNSVDMCFY	241
QY	241	DGYSTNSSSLEIRPDNNPNKSDGK	FYLKINDDTKRIAVTSLILLAGKSLTPNGTSLNI	300
Db	242	DGYSTNSSSLEIRPDNNPNKSDGK	FYLKINDDTKRIAVTSLILLAGKSLTPNGTSLNI	300
QY	301	ADAASLEFNMMNRITAVTMEPISV	PVLCWPGRLQDLDAKVENPEAGQYMGNIINVTFTSSQOT	360
Db	301	ADAASLEFNMMNRITAVTMEPISV	PVLCWPGRLQDLDAKVENPEAGQYMGNIINVTFTSSQOT	360
QY	361	L 361		
Db	361	L 361		

```

RESULT 3
US-09-839-894-29
; Sequence 29, Application US/09839894
; Patent No. US20020176868A1
; GENERAL INFORMATION:
; APPLICANT: Allboun, Zeev
; APPLICANT: Barry, Eileen M.
; APPLICANT: Levine, Myron M.
; APPLICANT: University of Maryland
; TITLE OF INVENTION: ISOLATION AND CHARACTERIZATION
; OF THE 3' UNTRANSLATED REGION OF THE 5' UNTRANSLATED
; FILE REFERENCE: USOMD, 0066A
; CURRENT APPLICATION NUMBER: US/09/839,894
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: 60/198,626
; PRIOR FILING DATE: 2000-04-20
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 359
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: ETEC Protein Homology Sequence
; US-09-839-894-29

```

Query Match	81.8%	Score 1542;	DB 9;	Length 359;
Best Local Similarity	80.9%	Pred. No. 3.3e-132;		
Matches 292: Conservative	32	Mismatches 152	Indels 2	Gaps 2

QY	1	MNLLFFLFFLFFSVLLTFFPAVSADKIRGDSITFNI	PGPRDRNNESSPKHNI	NNHTFAYSE	60
Dd	1	MNKLFLFFLFFLFFSSGFFTFPAVSADKNPGESEMI	NTTICPHRGGSSP	YNNLNSTLYANG	60
QY	61	SHLLYDRMTFLCLSSNHTLNLGACETSEPNSSSV	SGEFTNTTLOFTEKRS	LIKRELOIKGY	120
Dd	61	SHHLYDRMSFLCLSSQNTLNLGACPSPADG	SNATIDGETNTITLOFTEKRS	LIKRELOIKGY	120
QY	121	KQLLFKSYNCCSGGLTNSAHFNCKNNAASGAS	LSLYTIPAGELNKL	PRGGIWDATLKL	180
Dd	121	KQFLFKNANCPSKIALNLSHQCRCREASGAT	LSLYTIPAGELNKL	LPBGVWNAVLKL	180
QY	181	RRYSEGTGYVINITIKLTIDKGNIOQLPOFK	SARDLMLRPTGSGTY	IGRANSVDMCFY	240
Dd	181	RRYTYGTGYTINTVNLTDKGNIOQLPOFK	SNARVDLMLRPTGSGTY	IGRANSVDMCFY	239
QY	241	DGYSTNSSSLEIRFODNNPKSKDGFYLR	KINDTDEKELAYTL	SELLACKSLTPNGTSLNT	300
Dd	240	DGYSTMSSEIRFODNNSKSKDGFYLR	KINDDSKELAYTL	SELLACKSLTPNGTSLNT	299
QY	301	ADAASLETNNNRITAYVMPPELSVYVLCW	PGRLQDDAAVENP	PEAGQYVGNINVTFTSSQT	360
Dd	300	-NFTASLETNNNRITAYVMPPELSVYVLCW	PGRLQDDAAKVNPEAGQYVGNIKITFTPSQT	358	
QY	361	L	361		
Dd	359	L	359		

```

: RESULT 4
: US-09-839-894-31
: Sequence 31, Application US/09839894
: Patent No. US2002017668A1
: GENERAL INFORMATION:
: APPLICANT: Alboum, Zeev
: APPLICANT: Barry, Eileen M.
: APPLICANT: Levine, Myron M.
: TITLE OF INVENTION: University of Maryland
: TITLE OF INVENTION: ISOLATION AND CHARACTERIZATION OF THE
: FILE REFERENCE: USFMD.006A
: CURRENT APPLICATION NUMBER: US/09/839,894

```



```

OY      236 DMCYYDGTSSSELETRFOONNPKSDGKFLRKINDOTREIAVTLSELLAGKSLPPTNG 295
        ||| ||| : ||| : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      234 DMLCYDG- GYKARSLQW--MEGSUKSCTGRQYIK-SDSADTIDXAAYSMMNYGGRSITPYTRG 289
OY      296 TSLNIADAASLETWMNRRTAVTAVTMPETISVPVLCWPGRQLDLAK---VENPEAQGYMINIV 352
        ::::| :||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db      290 VEFSLIDNVDKAATR-----PVALTGQRQAQVACVPLLTTPQPNIREKRSGEGYGTLLTV 344
OY      353 TEFPSSQT 360
        ||: ||
Db      345 TMLMGTOI 352

RESULT 7
US-09-303-232-2
: Sequence 2, Application US/09303232A
: Patent No. US20020006657AI
: GENERAL INFORMATION:
: APPLICANT: Bayer Aktiengesellschaft
: TITLE OF INVENTION: Nucleic Acids which encode
: TITLE OF INVENTION: Insect acetylcholine receptor subunits
: FILE REFERENCE: I.e A 33 020-Foreign Countries
: CURRENT APPLICATION NUMBER: US/09/303.232A
: CURRENT FILING DATE: 1999-04-30
: EARLIER APPLICATION NUMBER: DE 198 19 829.9
: EARLIER FILING DATE: 1998-05-04
: NUMBER OF SEO ID NOS: 6
: SOFTWARE: Patentin Ver. 2.1
: SEQ ID NO 2
: LENGTH: 770
: TYPE: PRNT
: ORGANISM: Drosophila melanogaster
: US-09-303-232-2
```

```

Query Match 5.3%; Score 119.5; DB 10; Length 770;
Best Local Similarity 20.4%; Pred. No. 0.0083;
Matches 83; Conservative 48; Mismatches 140; Indels 135; Gaps 20

QY      9 TLFSSVLTFAVSAADKIPGDESTINIFGPRDRNESSPKHNLINNH-----54
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      249 TIAFISYIGSFAOQLKNSSSSSSSN-----SSNSSSTQILNKLKHSWIFLLYLNLSAK 304
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      55 --ITAVSESHLYDWMTEFLCLSSNHTLNGACPTSNPSSSVSGEINITLOTEKRSLIK 112
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      305 VCLAGYHEKRLHND-----LDPPYNTLERPVYNESPOLLSGLTLMQIIDVDEKQQLLV 359
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      113 RELQIKGYKQKQLLFKSVNGPSGLTILNSAHENCNKNAASGASLYLIPAGELKLNLPFGIMD 172
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      360 TNVMWLK-----LEWMDMLNRWMTSDYGVK-----DLR-IPPRIRMK 395
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      173 ATLKLRVKKRRSET-----YGYTTINTITKLDKGNIQIWLPO--FKSDARVDLNLRPFG 226
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      396 PDVILM-----YNSADEGEGFYQTNNVVR--RNGSC-LYVPGIIFKSTCKIDITWPP-- 444
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      227 GYIGRNSVDKMF-----YDGYSTNSSSLERODNNPKS-----DGKYLKRINDOTKE 276
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      445 ---FDDQRCMKFGSWTIDGF-----QLDQLQDDEIGGISSTVYLNGEWELLGVPCKRNE 496
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      277 IAYT-----LSLLAGKSLTPTNGTSLNIADA 303
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      497 IYVNCCEPEYIDITFAIIIRRTLYEFFNLIPCVLIASMALLGFTLPDSCGKSLIGVT 556
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      304 ASLE-TNNNRITAVTMPETS--VPV-----LCW-----PGR 331
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      557 ILLSLTVFLNVAETMPATSDAVPLMIRIVFLCWLPLMILMSRPGR 602
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 8
US-10-092-880-10
; Sequence 10, Application US/10092880
; Patent No. US20020164354A1
; GENERAL INFORMATION:

```

```

; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS OF NON-TYPEABLE
; TITLE OF INVENTION: HAEMOPHILUS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/092,880
; CURRENT FILING DATE: 2002-03-08
; PRIOR APPLICATION NUMBER: 09/155,614
; PRIOR FILING DATE: 1998-09-30
; PRIOR APPLICATION NUMBER: 08/617,697
; PRIOR FILING DATE: 1996-04-01
; PRIOR APPLICATION NUMBER: PCT/US97/04707
; PRIOR FILING DATE: 1997-04-01
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 1600
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
; US-10-092-880-10

```

```

Query Match          6.2%; Score 116.5; DB 9; Length 1600;
Best Local Similarity 21.6%; Pred. No. 0.044;
Matches 74; Conservative 60; Mismatches 130; Indels 79; Gaps 19;

QY      48  HNINNHITATYSESHITLYDRMTFLCLSSHNTLNGACPTSENPSSS----SVSGETNTTL 102
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      816  YNEXSKH--ALNNSHNL-----TILGNAVTLGC-----ENSSSTITGNININTKANVTL 862

QY      103  QP-----TEKRLKIRELOIKGYKOLFKRSVNCPSGTLT-----NSAHFQCN--- 144
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      863  QADTNSNSTGLKRFLLTLGNISVEGNLSLGTGANANIVGNLSIAEDSTERGEASDMLNTTG 922

QY      145  ---KMAAGCASLTYLIPAGEELKNLPFGIMDATELRYKRRYSETYGT--YTANITIKLTD 200
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      923  TFTNNGTANINIKGVYKGLDINK--GG-----LNTTANSOGTKILININTI---NE 970

QY      201  KGNIDWLPQKRSQDARVDL--NLRPYGGGYTIGRNSVDMCFYDGYSTNSSSLEIRFQ--- 255
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      971  KGLDNI--KNIKADAEIOIGNISOEGNLTFTSSDKVNI-----TNOITKAGVEGR 1021

QY      256  -DNNKPSGGRKYLRKINDTKEIATLISLLAG---KSITPTNGTSLNITADAASLETNNN 311
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      1022  SSSSEAEVANNTIQ-----TEKLKLAGDLNITSGFKAEITANGSDLTIGNASGNADAK 1076

QY      312  RITAVTMEPISVPLYLCPMG-PRLQDAKYENPEAGYMGYNINYT 353
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      1077  K--VTFDKVKMDSKISTDGHNVTLINSEVKTSIGSSNAGNDNST 1116

RESULT 9
US-09-912-020-302
: Sequence 302, Application US/09912020
: Patent No. US20020045592A1
: GENERAL INFORMATION:
: APPLICANT: Zyskind, Judith
: APPLICANT: Ohlsen, Kari L.
: APPLICANT: Trawick, John
: APPLICANT: Forsyth, R. Allen
: APPLICANT: Fireolich, Jamie M.
: APPLICANT: Carr, Grant J.
: APPLICANT: Yamamoto, Robert T.
: APPLICANT: Xu, H. Howard
: TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
: TITLE OF INVENTION: ESCHERICHIA COLI
: FILE REFERENCE: ELITRA 001DAVI
: CURRENT APPLICATION NUMBER: US/09/912.020
: CURRENT FILING DATE: 2001-07-23
: PRIOR APPLICATION NUMBER: 09/492,709
: PRIOR FILING DATE: 2000-01-27
: PRIOR APPLICATION NUMBER: 60/117,405
: PRIOR FILING DATE: 1999-01-27
: NUMBER OF SEQ ID NOS: 485
: SOFTWARE: FastSeq for Windows Version 3.0

```



SEQ ID NO 302  
LENGTH: 2383  
TYPE: PRT  
ORGANISM: E. COLI  
US-09-912-020-302

Query Match 5.8%; Score 109; DB 10; Length 2383;  
Best Local Similarity 20.0%; Pred. No. 0.37;  
Matches 78; Conservative 65; Mismatches 157; Indels 90; Gaps 15;

QY 13 SSVLFPAVSADKIPDEDTITNFGPRDRNESSPKHILNNHTAY---SESHLTDRMT 69  
DB 762 SAKIATLASNGVLANEAMNTVSVNADEGS---NPINDHTVTPAVLGSATSPNNON 818  
QY 70 FLCLSSHTNLNGACPTSENPSSSVSGETNITLTOTFEKRSLLI-----K 112  
DB 819 ----TAKTIVNGCLA-TFDLKSSSQEDNTVEVTLENGVKOTLIYSPGDSSTAQVLDKSK 873  
QY 113 RELQIKRYQL-LFKSVNCPGSLTILNSAHFNCKNAASGLYLIYPAGELKNLPFGGIW 171  
DB 874 NEVVADGNDVSMTATVRDAKGNLNDVAVTFVNSAEAKLSQTEVNSHD-----GIA 926  
QY 172 DATLKLARKVRRSEYTGTTITIKLTDKNIQIWLPOPKSDARVDLRLPTGGCTYIG 231  
DB 927 TATLT-----SLKNGDYRYTASVSSGSOANOQVNFIDOSTAALTLVS-PSGDITV-- 976  
QY 232 RNSVDMCFYDGYSTNSS-----SLEIRFODNNPKSDGKFLRKINDTKELIAYTILLSLAG 287  
DB 977 -----TNTAPQYMTATLTDKNGNPLKDKETFFSPND-----VASKFSISNGG 1019  
QY 288 KSLTPTNGTSL-----NIADAASLETMMNRTTAVTMEPISVPLC 327  
DB 1020 KMTDSNGVAIASLTGTLAGTHIMARLANSVSDAQPMTFVADKRAVVVLOTSKAEII 1079  
QY 328 WPG--RLQDAKVENPEAGQMGNIWTF 355  
DB 1080 GNGVDETTLTATVKDP-SNHPVAGITVNF 1108

RESULT 10  
US-10-092-880-9

; Sequence 9, Application US/10092880  
; Patent No. US20020164354A1  
; GENERAL INFORMATION:  
; APPLICANT: Barenkamp, Stephen J.  
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS OF NON-TYPEABLE  
; TITLE OF INVENTION: HAEMOPHILUS

FILE REFERENCE:  
CURRENT APPLICATION NUMBER: US/10/092,880  
CURRENT FILING DATE: 2002-03-08  
PRIOR APPLICATION NUMBER: 09/155,614  
PRIOR FILING DATE: 1998-09-30  
PRIOR APPLICATION NUMBER: 08/617,697  
PRIOR FILING DATE: 1996-04-01  
PRIOR APPLICATION NUMBER: PCT/US97/04707  
PRIOR FILING DATE: 1997-04-01  
NUMBER OF SEQ ID NOS: 11  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 9  
LENGTH: 1599  
TYPE: PRT  
ORGANISM: Haemophilus influenzae  
US-10-092-880-9

Query Match 5.4%; Score 101; DB 9; Length 1599;  
Best Local Similarity 21.5%; Pred. No. 1.1;  
Matches 90; Conservative 52; Mismatches 172; Indels 104; Gaps 18;

QY 21 VSADKIPGDESTINFGPRDRNESSPKHNI-LNNHT-TAYSSHLYDRTFLCLSS---- 74  
DB\* 1077 VTFDKV-----KDSKISTDGHNVTLNSEKVTGSSNAGNDNSTGLTISAKDV 1124  
QY 75 -----SHNTLN-----GACPTSENPSSSVSGETNITLTQ-----FTEKR 108

DB 1125 TVNNVTSKHTINISAAAGNVTTKEGTTINATGVSVEVTAQNTIKNTISQNVYTAPE 1184  
QY 109 SLIKRELQI-----KGYKOLLFKSVNCPGSLTILNSAHFNCKNAASGLI----- 153  
DB 1185 NLTTEENAVINATSGTVNSTKGTGDKIGIEESTSGNVNI---TASGNTLKVSNITGQDVT 1241  
QY 154 -----LYIIPAGELKNLPFGGIWMDATLKLARKVRRSEYTGTTITIKLTDKNGIQIW 207  
DB 1242 VYADAGALTTTGSTISATGANNATTTKTGDINGKVESSSGSVTLTATATLAVNGISGN 1301  
QY 208 LPOPKSDARVDLRLPTGGCTYIGRNSVDMCFYDGYSTNSSSLEIRFODNN---PKSDGK 264  
DB 1302 TVTTFADSG---KLSTGVSTINGTNSVTT-----SSQSGDIEGTISGNTVNVTAASGD 1352  
QY 265 FYL-RKINDTEIAYTLS-----LLAGKSLTPNGTSLIADAASLETMMN--RITA 315  
DB 1353 LTIGNSAKVEARNGAATTLAESGKLTPTTGSSITSSNGTTLTARDSSIAGININANAYTL 1412  
QY 316 VTMPEISV-----PVLCPGRQLQ---DAKVENPEAG-----QYMGNIWTFPPSS 358  
DB 1413 NTTGTLTTTGDSKINATSGTLTINAKDAKLDGASGDRIVNATNAGSGNVTAKTSS 1470

RESULT 11  
US-09-839-894-6

; Sequence 6, Application US/09839894  
; Patent No. US20020176868A1  
; GENERAL INFORMATION:  
; APPLICANT: Alboum, Zeev  
; APPLICANT: Barry, Eileen M.  
; APPLICANT: Levine, Myron M.  
; TITLE OF INVENTION: ISOLATION AND CHARACTERIZATION OF THE  
; TITLE OF INVENTION: CSA OPERON  
; FILE REFERENCE: USFMD.006A  
CURRENT APPLICATION NUMBER: US/09/839,894  
CURRENT FILING DATE: 2001-04-20  
PRIOR APPLICATION NUMBER: 60/198,626  
PRIOR FILING DATE: 2000-04-20  
NUMBER OF SEQ ID NOS: 40  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 6  
LENGTH: 867  
TYPE: PRT  
ORGANISM: E. COLI  
US-09-839-894-6

Query Match 5.0%; Score 95; DB 9; Length 867;  
Best Local Similarity 23.6%; Pred. No. 1.6;  
Matches 47; Conservative 24; Mismatches 56; Indels 72; Gaps 10;

QY 125 FKSVCNCPGSLTILNSAH-----FNCKNAASGASLYIYPAGELKNLPFGCIWDATLKLRY 179  
DB 435 YOSISTYDGFSLSFYHNDKRVNDCGRNYNAGWS-GCYESYSASLSIPLIG-WTSTL----- 488  
QY 180 KRYSET-----YGFYINITIKLTDKNGIQIW-LPOPKSDARVDLRLP-- 223  
DB 489 --GYSDTISESYKKNHIISEYGFYNONTI---KGRTORWOLTSSTLSKMDVYFMPI 541  
QY 224 -----TGCGTYI-----GNSVDMCFYDGYSTNSSS-- 249  
DB 542 GIYNSRQRLTLDKGVISVTLTRASHRENSLNGYSVYNSRKYSSNLFVDDGYMTSTNGG 601  
QY 250 -----LEIRFODNNPKSDGK 264  
DB 602 DYHEVAMREKKNRHNNAEGR 620

RESULT 12  
US-09-797-862-33  
; Sequence 33, Application US/09797862  
; Patent No. US20020102276A1

Query Match	5.08;	Score 94.5;	DB 10;	Length 2353;
Best Local Similarity	19.78;	Pred. NO. 7.6;		
Matches 61;	Conservative 47;	Mismatches 117;	Indels 85;	Gaps 14;

```

: U/LT 13
: 09-924-154-14
: Sequence 14, Application US/09924154
: Patent No. US20020127241A1
: GENERAL INFORMATION:
: APPLICANT: Narum, David L.
: APPLICANT: Sim, Kim L.
: TITLE OF INVENTION: Anti-Plasmodium Compositions and Methods of Use
: FILE REFERENCE: 05213-0465 43170-262105
: CURRENT APPLICATION NUMBER: US/09/924,154
: CURRENT FILING DATE: 2001-08-07
: PRIOR APPLICATION NUMBER: US 60/223,525
: PRIOR FILING DATE: 2000-08-07
: NUMBER OF SEQ ID NOS: 17
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 14
: LENGTH: 1143
: TYPE: PRT
: ORGANISM: Mammalian
: US-09-924-154-14

```

```

RESULT 14
US-09-944-160-12
; Sequence 12, Application US/09944160
Patent No. US20020174452A1
GENERAL INFORMATION:
APPLICANT: Lewis, No. US20020174452A1man
APPLICANT: Davis, Laurence
APPLICANT: .. Huang, Ming
TITLE OF INVENTION: Monocot Seeds with Increased Lignan
TITLE OF INVENTION: Content
FILE REFERENCE: WSRU117983
CURRENT APPLICATION NUMBER: US/09/944,160
CURRENT FILING DATE: 2001-08-30
PRIOR APPLICATION NUMBER: US 60/2230,632
PRIOR FILING DATE: 2000-09-07
NUMBER OF SEQ ID NOS: 49
SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 564
; TYPE: PRF
; ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: lacase amino acid sequence from plasmid pAF1245
US-09-944-160-12

```

	Query Match	4.88;	Score 90.5;	DB 9;	Length 564;
	Best Local Similarity	22.08;	Pred. No. 2.3;		
	Matches	80;	Conservative	57;	Mismatches 132; Indels 95; Gaps
OY	1	MKKLIFT-----FLTFSSVLPTFAVSADKI---	PDDESTINTFFGPRDRESSPKINIL	51	
		:   :	:   :		
Db	230	MNTLIESFIANHSYTVAGSDAAVYKPLSKSDITTPSG-QTIDFLL----	QANDQPSHYIM	284	
OY	52	NNHTAVASESHTELYDRMTFLCLSSHNTLNGACPTSENPSSSVSG--ETNITLOFTEK-R	108		
		:   :	:   :		
Db	285	AARAAVAAGN---FDNTTTTTAIIIRKG-NYPAPES--PSPFNLECGFDNTMNAVVFYYRLR	336		
OY	109	SILIKRELDIGKYK---OLLRK-SVN--CPESGLTNSAHNCNKKNNAASGASYLYIPAG	166		
		:   :	:   :		
Db	339	SLSGNKYVDVPKPNVTDKLTTLTFISNLTPCEN----NSCAAPFERRRASVNNITTFP-	392		
OY	161	EKLNFPGCIDADATLKLVKRYKRYSETYT---YTNITTIKTLDKGNIOIWLPOKSDARV	217		
		:   :	:   :		

Db 333 -----PTAILOAYQ-RIRNVYSNNPSPFPFENYTSIIIPR---DLMRPQNGTEVKV 443  
 Oy 218 DLNLAPRTGGGYIGNSVDKCFYDGYSTNSSLEIRPDNNPKSGKRYLKRINDTKEI 277  
 Db 444 -----LKNSTVEIVFOGTN-----ILAGIDH----- 465  
 Oy 278 AVTLSLIAGKSLPTNGTSLINADASLETNMNRITAVTMEISVPVLCMPGRLOLDAK 337  
 Db 466 -----PIHLHGOSFYV-CWGLGNFNNAIDPLNLYNDPPLNNTIAPVSGWTA---VRFK 517  
 Oy 338 VEMP 341  
 Db 518 ASNP 521

RESULT 15

US-10-011-588-29  
 ; Sequence 29, Application US/10011588  
 ; Patent No. US20020168727A1  
 GENERAL INFORMATION:  
 APPLICANT: Smith, Leonard  
 ; TITLE OF INVENTION: RECOMBINANT LIGHT CHAINS OF BOTULINUM  
 ; TITLE OF INVENTION: NEUROTOXINS AND LIGHT CHAIN FUSION PROTEINS FOR USE IN  
 ; TITLE OF INVENTION: RESEARCH AND CLINICAL THERAPY  
 ; FILE REFERENCE: A34796 067252.013  
 ; CURRENT APPLICATION NUMBER: US/10/011,588  
 ; CURRENT FILING DATE: 2002-03-29  
 ; PRIOR APPLICATION NUMBER: 09/910,186  
 ; PRIOR FILING DATE: 2001-07-20  
 ; PRIOR APPLICATION NUMBER: 09/611,419  
 ; PRIOR FILING DATE: 2000-07-06  
 ; PRIOR APPLICATION NUMBER: 60/246,744  
 ; PRIOR FILING DATE: 2000-11-06  
 ; PRIOR APPLICATION NUMBER: 60/311,966  
 ; NUMBER OF SEQ ID NOS: 47  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 29  
 ; LENGTH: 811  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Recombinant protein encoded by SEQ ID NO:28  
 US-10-011-588-29

Query Match 4.7%; Score 89.5; DB 9; Length 811;

Best Local Similarity 18.4%; Pred. No. 4.7;  
 Matches 76; Conservative 60; Mismatches 112; Indels 165; Gaps 19;

Oy 2 NKILFIFLFFSSVLETPFAVSADKIPGDESTITFGPRDRNESSPKHNLNNHITAYSES 61  
 Db 351 NELQIFTEFYAKIYN--VONRKI---YLSNYTP-----VTANILD 388  
 Oy 62 HTLYDRMTFEL--SSHTLNGACPTSENPSSSVSGETNITLOFTEKRSILKRELQIKG 119  
 Db 389 DNAYDIQNGFNIPKSNLVLTMGONLSRNPARKVNP-NMLYLFKFCSLNKTLDG-- 445  
 Oy 120 YKQLFKSVNCPGSLTNSAHFNCNKMAASGLYIIPAGELKNLPF-GGIWDATLKLK 178  
 Db 446 -RELVKNT-----DLPIGDISDYKTDIF 469  
 Oy 179 VKRRYSEYGT--YTINITI-----KLTDKNIQIWLPOFKSPDARYDLNLPTGGGTY 229  
 Db 470 LKRDINEETEVLYYPDNVSVDVILSKNTSEHGQDLILPSIDSESEI-----L 518  
 Oy 230 IGRNSVDKCFYDGYSTNSSLEIRPDNNPKSGKRYL--RKINDTKEIAYTLSTL-- 283  
 Db 519 PGENOV---FTDNKIQNDYIL-----NSYYLESQKLSDNVEDFTTTRISIERAL 564  
 Oy 284 -----LLAGK-----SLTPTN 294  
 Db 565 DNSAKVYTYTFPLANKVNAGVGGFLMAMANDVVEDFTTNILRKDTLDKISDVSAITPYI 624

Oy 295 GNSLNIADAASLETNMNRITAVT-----MPETSVPLCMPSGRLOLDAKVE 339  
 Db 625 GPALNINSVR-RGNFTEAFAYGVITILEAFPEFTIPAL--GAFVYSKVQ 673

Search completed: December 6, 2002, 18:19:11  
 Job time : 18 secs

**This Page Blank (uspto)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 6, 2002, 17:55:05 ; Search time 17 Seconds  
(without alignments)  
624.804 Million cell updates/sec

Title: US-09-839-894-10

Perfect score: 1886  
Sequence: 1 MNKILFIPTLFSSVLEFFFA.....EAGGYMGNINVTTPSSQTL 361

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

tal number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PTUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1536.5	81.5	360	2	US-08-483-101-17 Sequence 17, Appl
2	988.5	52.4	363	2	US-08-483-101-16 Sequence 16, Appl
3	927.5	49.2	364	2	US-08-483-101-5 Sequence 5, Appl
4	116.5	6.2	1600	2	US-08-617-697-10 Sequence 10, Appl
5	114	6.0	650	3	US-08-362-525-2 Sequence 2, Appl
6	111.5	5.9	1005	4	US-09-206-942-41 Sequence 41, Appl
7	111.5	5.9	1011	4	US-09-206-942-39 Sequence 39, Appl
8	111.5	5.9	1529	2	US-08-728-470-10 Sequence 10, Appl
9	111.5	5.9	1529	4	US-08-719-641-10 Sequence 10, Appl
10	107	5.7	1095	4	US-09-206-942-45 Sequence 45, Appl
11	107	5.7	1101	4	US-09-206-942-43 Sequence 43, Appl
12	106	5.6	671	2	US-08-737-716-13 Sequence 13, Appl
13	101	5.4	915	4	US-09-206-942-35 Sequence 35, Appl
14	101	5.4	1222	4	US-09-206-942-37 Sequence 37, Appl
15	101	5.4	1228	4	US-09-206-942-34 Sequence 34, Appl
16	101	5.4	1338	2	US-08-728-470-9 Sequence 9, Appl
17	101	5.4	1338	4	US-08-719-641-9 Sequence 9, Appl
18	101	5.4	1599	2	US-08-617-697-9 Sequence 9, Appl
19	100.5	5.3	1004	4	US-09-206-942-57 Sequence 57, Appl
20	100.5	5.3	1010	4	US-09-206-942-55 Sequence 55, Appl
21	99.5	5.3	969	4	US-09-206-942-32 Sequence 32, Appl
22	99.5	5.3	975	4	US-09-206-942-30 Sequence 30, Appl
23	99	5.2	992	4	US-09-206-942-61 Sequence 61, Appl
24	99	5.2	998	4	US-09-206-942-59 Sequence 59, Appl
25	99	5.2	1848	4	US-08-296-791-6 Sequence 6, Appl
26	99	5.2	1848	5	PCT-US95-10661A-6 Sequence 6, Appl
27	97	5.1	977	4	US-09-206-942-53 Sequence 53, Appl

28	97	5.1	983	4	US-09-206-942-51 Sequence 51, Appl
29	94.5	5.0	1912	1	US-08-409-995-4 Sequence 4, Appl
30	94.5	5.0	1912	3	US-08-685-467-4 Sequence 4, Appl
31	94.5	5.0	2353	4	US-09-377-155-33 Sequence 33, Appl
32	94.5	5.0	2353	4	US-08-913-942-4 Sequence 4, Appl
33	94.5	5.0	2353	4	US-09-669-974-33 Sequence 33, Appl
34	94.5	5.0	2354	4	US-09-268-347-47 Sequence 47, Appl
35	94.5	5.0	2411	4	US-09-268-347-36 Sequence 36, Appl
36	93.5	5.0	901	4	US-09-134-001C-5351 Sequence 5351, Ap
37	93	4.9	1167	1	US-08-485-568A-6 Sequence 6, Appl
38	93	4.9	1167	2	US-08-590-554A-6 Sequence 6, Appl
39	93	4.9	1167	2	US-09-184-223-6 Sequence 6, Appl
40	92.5	4.9	1073	4	US-09-206-942-49 Sequence 49, Appl
41	92.5	4.9	1079	4	US-09-206-942-47 Sequence 47, Appl
42	91.5	4.9	869	2	US-08-483-101-15 Sequence 15, Appl
43	91	4.8	1168	1	US-08-620-717A-9 Sequence 9, Appl
44	90.5	4.8	1328	3	US-08-781-891-76 Sequence 76, Appl
45	90	4.8	569	4	US-09-173-053-18 Sequence 18, Appl

#### ALIGNMENTS

```
RESULT 1
US-08-483-101-17
; Sequence 17, Application US/08483101
; Patent No. 5932715
; GENERAL INFORMATION:
; APPLICANT: Scott, June R.
; APPLICANT: Froehlich, Barbara
; TITLE OF INVENTION: CS2 Proteins and Coding Sequences
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Greenlee and Winner, P.C.
; STREET: 5370 Manhattan Circle, suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: US
; ZIP: 80503
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,101
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Feiber, Donna M.
; REGISTRATION NUMBER: 33878
; REFERENCE/DOCKET NUMBER: 6-95
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 499-8080
; TELEFAX: (303) 499-8089
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 360 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; US-08-483-101-17

Query Match      81.5%; Score 1536.5; DB 2; Length 360;
Best Local Similarity 80.1%; Pred. No. 8e-150;
Matches 289; Conservative 33; Mismatches 38; Indels 1; Gaps 1;
```

QY 1 MNKILFIPTLFSSVLEFFFAVSADKIPGDESTNIFGPRDNSSPKHINLNHTAYSE 60  
DB 1 MNKILFIPTLFSSGFFFAVSADKNPGSENMNTIGPHDRGSSPIYINLSYLTAYNG 60

Qy	61	SHTLYDRMFTCLSSHNTLNACPTSENPSSVSGENTLLOPTEKRSLLIKRELQJG	120
Dp	61	SHHLYDRMSFCLSSQNTLNACPPSDAPGATIDGETINTILOPTEKRSLLIKRELQJG	120
Qy	121	KOLLEKSVNCPSGTLNLSAHFENCKNNAASGASLYLYIPAGEKLNLPFGIWDATLKIrvk	180
Dp	121	KOLPEKMNACPSKIALNLSNHRQCRREOASGATLSLYIPAGELNMLPFGVWNVLLKIrvk	180
Qy	181	RRYSEYCTYVINTIKTLTDKGNIOIWLPOKSPARVDNLNRPFGGTYIGRNSVDMCFY	240
Dp	181	RRYDTYCTYVINTIYNLTLDKGNIOIWLPOKSNARVDNLNRPFGGTYIGRNSVDMCFY	240
Qy	241	DGYSTNSSLLEIRQDNNPKSGDFYLRKINDDTKEIAYTLSSLACKSLTPPTNGISLNT	300
Dp	241	DGYSTNSSLLEIRQDNNKSKDGKFYLRKINDDSKEIAYTLSSLACKNLTPTNGALNT	300
Qy	301	ADAASLEFNMRITAVTPEISVYVLCWPGRLQJDAKVENDEAGQYNGINVTFFPSQT	360
Dp	301	NTASLEFNMRITAVTPEISVYVLCWPGRLQJDAKVENDEAGQYNGINVTFFPSSSS	359
	361	L 361	
Dp	360	L 360	

```

1      RESULT 2
2      US-08-483-101-16
3      ; Sequence 16, Application US/08483101
4      ; Patent No. 5932715
5      ; GENERAL INFORMATION:
6      ; APPLICANT: Scott, June R.
7      ; APPLICANT: Froehlich, Barbara
8      ; APPLICANT: Caron, Judy
9      ; TITLE OF INVENTION: CS2 Proteins and Coding Sequences
10     ; NUMBER OF SEQUENCES: 17
11     ; CORRESPONDENCE ADDRESS:
12     ; ADDRESSEE: Greenlee and Winner, P.C.
13     ; STREET: 5370 Manhattan Circle, Suite 201
14     ; CITY: Boulder
15     ; STATE: Colorado
16     ; COUNTRY: US
17     ; ZIP: 80303
18     ; COMPUTER READABLE FORM:
19     ; MEDIUM TYPE: Floppy disk
20     ; COMPUTER: IBM PC compatible
21     ; OPERATING SYSTEM: PC-DOS/MS-DOS
22     ; SOFTWARE: PatentIn Release #1.0, Version #1.30
23     ; CURRENT APPLICATION DATA:
24     ; APPLICATION NUMBER: US/08/483,101
25     ; FILING DATE: 07-JUN-1995
26     ; CLASSIFICATION: 424
27     ; ATTORNEY/AGENT INFORMATION:
28     ; NAME: Feider, Donna M.
29     ; REGISTRATION NUMBER: 33878
30     ; REFERENCE/DOCKET NUMBER: 6-95
31     ; TELECOMMUNICATION INFORMATION:
32     ; TELEPHONE: (303) 499-8080
33     ; TELEFAX: (303) 499-8089
34     ; INFORMATION FOR SEQ ID NO: 16:
35     ; SEQUENCE CHARACTERISTICS:
36     ; LENGTH: 363 amino acids
37     ; TYPE: amino acid
38     ; STRANDEDNESS:
39     ; TOPOLOGY: unknown
40     ; MOLECULE TYPE: protein
41     ; HYPOTHETICAL: NO
42     ; US-08-483-101-16

```

Query Match	52.4%;	Score 988.5;	DB 2;	Length 363;
Best Local Similarity	55.0%;	Pred. No. 2.2e-93;		
Matches 204; Conservative	53;	Mismatches 93;	Indels 21;	Gaps 11;

[illegible]

```

1      RESULT 3
2      US-08-483-101-5
3      ; Sequence 5, Application US/08483101
4      ; Patent No. 5932715
5      ; GENERAL INFORMATION:
6      ; APPLICANT: Scott, June R.
7      ; APPLICANT: Froehlich, Barbara
8      ; APPLICANT: Caron, Judy
9      ; TITLE OF INVENTION: C52 Proteins and Coding Sequences
10     ; NUMBER OF SEQUENCES: 17
11     ; CORRESPONDENCE ADDRESS:
12     ; ADDRESSEE: Greenlee and Winner, P.C.
13     ; STREET: 5370 Manhattan Circle, Suite 201
14     ; CITY: Boulder
15     ; STATE: Colorado
16     ; COUNTRY: US
17     ; ZIP: 80303
18
19     ; COMPUTER READABLE FORM:
20     ; MEDIUM TYPE: Floppy disk
21     ; COMPUTER: IBM PC compatible
22     ; OPERATING SYSTEM: PC-DOS/MS-DOS
23     ; SOFTWARE: Patent In Release #1.0, Version #1.30
24     ; CURRENT APPLICATION DATA:
25     ; APPLICATION NUMBER: US/08/483,101
26     ; FILING DATE: 07-JUN-1995
27     ; CLASSIFICATION: 424
28     ; ATTORNEY/AGENT INFORMATION:
29     ; NAME: Ferner, Donna M.
30     ; REGISTRATION NUMBER: 33878
31     ; REFERENCE/DOCKET NUMBER: 6-95
32     ; TELECOMMUNICATION INFORMATION:
33     ; TELEPHONE: (303) 499-8080
34     ; TELEFAX: (303) 499-8089
35     ; INFORMATION FOR SEQ ID NO: 5:
36     ; SEQUENCE CHARACTERISTICS:
37     ; LENGTH: 364 amino acids
38     ; TYPE: amino acid
39     ; TOPOLOGY: linear
40     ; MOLECULE TYPE: protein
41
42     ; US-08-483-101-5

```

Query Match	49.28;	Score 927.5;	DB 2,	Length 364;
Best Local Similarity	48.18;	Pred. No. 4.3e-87;		
Matches 176; Conservative	65;	Mismatches 118;	Indels 7;	Gaps 3

OY 1 MNKILFFLEFSSVLTFAVSADKIPGDESTINIFGPRDRN--ESSPKHNLNNHTAY 58  
1 LKKVIFVLSMFLCSQVYQSGMHTNVEAGSINKTESIGIPDRSAASYPAHYTFHEHVACY 60  
OY 59 SESHLYDRMFLCLSSHNTLNGACPTSENPSSSVSGETINTTLOFTEKRSILKELOJIK 118  
61 NNDHSLFMRMFLCMSSDASGKACPTGENSKS--QGFTNIKILFTEKKSILARLTNLK 118  
OY 119 GKOLLEFSSVNC---PSGLTLNSAFNCKNKAASGASLYLYIPAGELKLPFGGIMDATL 175  
119 GKRLFLYSDRCIHVYDKNNLSHFKCVGSFTRGVDFLLYPOGEIDOLLGGIWEAL 178  
OY 176 KLRVRRSEFETTYTINTIKLTDKGNIOIWLPOFKSPARDLNLRLPTGGSTYIGRNSV 235  
179 ELRVRRHNDYNGITKAVNTYDLTDKGNIOVTPKFSHPRIDLNLREPKNGKYSGSNVL 238  
OY 236 DMCIFYDGYSTNSSLFIRQDNNPKSDGKFYLRLKINDTKEIAYTLISLLAGSKLPTNG 295  
239 EMCLYDGYSTHQSISLMEFRQDSQGTGNNEXNLIKTGEPLKLPYKLSLLGGREFYPNG 298  
OY 236 TSLNLTADALETNNRITAYTAPPEISVPLVLCWPGRLQDAVENDPEAGQYMGINVTFT 355  
239 EAFITNDSSLEFINNRKISVSLPOISIPVLCWPAFLTMSLNPDEAGEYSGLINVTFT 358  
OY 356 PSSOTL 361  
OY 359 PSSSSL 364

RESULT 4  
US-08-617-697-10  
Sequence 10, Application us/08617697  
Patent No. 5977336

GENERAL INFORMATION:  
APPLICANT: Barenkamp, Stephen J  
TITLE OF INVENTION: High Molecular Weight Surface Proteins  
TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Shoemaker and Maltare, Ltd.  
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202-0286  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/617,697  
FILING DATE: 01-APR-1996  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/302,832  
FILING DATE: 05-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US PCT/US93/02166  
FILING DATE: 16-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Berkstresser, Jerry W  
REGISTRATION NUMBER: 22,651  
REFERENCE/DOCKET NUMBER: 10388-557  
TELEPHONE: (703) 415-0810  
TELEFAX: (703) 415-0813  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1600 amino acids  
TYPE: amino acid

STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-617-697-10

Query Match 6.2%; Score 116.5; DB 2; Length 1600;  
Best Local Similarity 21.6%; Pred. No. 0.016;  
Matches 74; Conservative 60; Mismatches 130; Indels 79; Gaps 19;

OY 48 HNLNNHTAYSESHLYDRMFLCLSSHNTLNGACPTSENPSSS---SVSGETINTL 102  
DB YNEYSKH--AINSSHNL-----TILGNAVTLGG-----ENSSSITGININTNNANYTL 862  
OY 103 GF-----TEKRSILKELOIKGYKOLFFRSVNCPSGLTL-----NSAHFNCN--- 144  
DB 863 QADTSNNTSGMLKRLTLTLGNISVEGNLSLTGNANAIYGLSLAEDSTPKEGASDLNLTG 922  
OY 145 ---KNAASGASLYLYIPAGELKLPFGGIMDATLKLVRKRRSEYGT-YTINITIKLTD 200  
DB 923 FETNNGTANINIKGVYKLGIDINNKK--GG-----LNTTNSAGTQKTIINGNIT---NE 970  
OY 201 KGNIOIWLPOFKSDARVDL--NLRLPTGGSTYIGRNSVDMCFYDGYSTNSSLFIRQ--- 255  
DB 971 KEDLNI--KNIKADAEIOIGNISQEGNLITRSSDKVNI-----TWOITKAGVEGGR 1021  
OY 256 -DNPKSDGKFYLRLKINDTKEIAYTLISLLAG---KSLPTNGTSLNLTADAASLETMNN 311  
DB 1022 SDSSFAENANLTIQ-----TKELKLAGDLNLSGFNKAEITTAANGSDLTIGNAGGNADAK 1076  
OY 312 RITAVTPEISVPLVLCWPG-RLOLDKAVENPEAGQYMGINVTFT 353  
DB 1077 K---VTFDKVKDSKISTDGHNVTLNLSVEKTSNGSSNAGDNST 1116

RESULT 5  
US-08-362-525-2  
Sequence 2, Application US/08362525  
Patent No. 6027910

GENERAL INFORMATION:  
APPLICANT: KLIS, FRANCISCUS M.  
APPLICANT: SCHREUDER, MARTEN P.  
APPLICANT: TOSCHKA, HOLGER Y.  
APPLICANT: VERRIPS, CORNELIS T.  
TITLE OF INVENTION: PROCESS FOR IMMOBILIZING ENZYMES TO THE  
TITLE OF INVENTION: CELL WALL OF A MICROBIAL CELL BY PRODUCING A FUSION  
NUMBER OF SEQUENCES: 26  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.  
STREET: 1100 New York Avenue, N.W.  
CITY: Washington  
STATE: D. C.  
COUNTRY: U.S.A.  
ZIP: 20005-3918  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/362,525  
FILING DATE: 04-JAN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: EP 92202080.5  
FILING DATE: 08-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: EP 92203899.7  
FILING DATE: 14-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/EP93/01763  
FILING DATE: 07-JUL-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: KOKULIS, PAUL N.





Oy 276 ELAYTLSELLAG---KSLPTNGTSLNIADASLETNNRITAVMPETISVPLCMGRL 332  
Db 569 ELLTLDNINISGPNKAEITADNSDLIGKASSDSNNAKOITFDVKXSKIS--AGNNIV 646  
Oy 333 OLDKVENPEA-----GQYMGNTVTFPPSSQTL 361  
Db 647 TLNSKVEITSNSDGTGSGSDNNIGLITISAKDVFV 681

## RESULT 8

US-08-728-470-10  
; Sequence 10, Application US/08728470  
; Patent No. 5928651  
; GENERAL INFORMATION:  
; APPLICANT: Barenkamp, Stephen J  
; TITLE OF INVENTION: High Molecular Weight Surface Proteins  
; TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Shoemaker and Mattare, Ltd.  
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
; STREET: Bldg. 1  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202-0286  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/728,470  
; FILING DATE:  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/302,832  
; FILING DATE: 16-MAR-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US PCT/US93/02166  
; FILING DATE: 16-MAR-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9205704.1  
; FILING DATE: 16-MAR-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Beirstresser, Jerry W  
; REGISTRATION NUMBER: 22,651  
; REFERENCE/DOCKET NUMBER: 1038-633  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 415-0810  
; TELEFAX: (703) 415-0813  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1529 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-728-470-10

Query Match 5.9%; Score 111.5; DB 2; Length 1529;  
Best Local Similarity 20.7%; Pred. No. 0.048;  
Matches 72; Conservative 62; Mismatches 127; Indels 87; Gaps 18;

Oy 48 HNIILNHTAASESTLYLDRMTFLCLSSHNLTNGACPTSEMPSS-----SVSGETNTTL 102  
Db 744 YNEYSKH--AINSSHNL-----TLIGGNVTLGG-----ENSSSSITGINITNANVTL 790  
Oy 103 QF-----TEKSLIKRELQIKGYKQLFKSVNCPGSLTILNSAHFNCNKNASGASLY 154  
Db 791 QADPSNMTGKRLTLTIGNISVEGNSLSLTGANNANIVGNLSI--AEDSTFGKAS----- 843  
Oy 155 LYIPAGELKNLPGGIW--DATLKLKVRKRRSEYGYT-----INITIKLT----- 199

Db 844 -----DNINIGTGTNNGTANINIKQGVYKLGIDINNKGLNITTNASGOTKTIING 895  
Oy 200 ----DKGNIQIWLPOFKDARVDL--NRPPTGGGYTIGRNSVDMKCFYGYSTINSSLEIR 253  
Db 896 NITNKGDLNT--KNIKADAETIGIGNISOKEGNLTJISSDKVNI-----TNQTTIAG 946  
Oy 254 FO----DNPNKSDGKFYRKINDPKEIAYTLSELLAG---KSLPTNGTSLNIADASL 306  
Db 947 VEGGRSDSSSEANANLTIQ-----TKELKLAGDLNISGFNAKEITAKKGSDLTITNASGG 1001  
Oy 307 ETNNRITAVTMPETISVPLCMG--RIOLDKVENPEAGQYMGNTV 353  
Db 1002 NADAKK---VTFDKVKDSKISTDGHNTVTLNSEVKTSGNSNAGDNST 1046

## RESULT 9

US-08-719-641-10  
; Sequence 10, Application US/08719641  
; Patent No. 6218141  
; GENERAL INFORMATION:  
; APPLICANT: Barenkamp, Stephen J  
; TITLE OF INVENTION: High Molecular Weight Surface Proteins  
; TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Shoemaker and Mattare, Ltd.  
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
; STREET: Bldg. 1  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202-0286  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/719,641  
; FILING DATE:  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/302,832  
; FILING DATE: 16-SEP-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US PCT/US93/02166  
; FILING DATE: 16-MAR-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9205704.1  
; FILING DATE: 16-MAR-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Beirstresser, Jerry W  
; REGISTRATION NUMBER: 22,651  
; REFERENCE/DOCKET NUMBER: 1038-625  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 415-0810  
; TELEFAX: (703) 415-0813  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1529 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-719-641-10

Query Match 5.9%; Score 111.5; DB 4; Length 1529;  
Best Local Similarity 20.7%; Pred. No. 0.048;  
Matches 72; Conservative 62; Mismatches 127; Indels 87; Gaps 18;

Oy 48 HNIILNHTAASESTLYLDRMTFLCLSSHNLTNGACPTSEMPSS-----SVSGETNTTL 102  
Db 744 YNEYSKH--AINSSHNL-----TLIGGNVTLGG-----ENSSSSITGINITNANVTL 790



```

: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCF/NL95/00170
: FILING DATE: 12-MAY-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: EP 94201353.3
: FILING DATE: 12-MAY-1994
: INFORMATION FOR SEQ ID NO: 13:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 671 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: ORIGINAL SOURCE:
: ORGANISM: Streptococcus faecalis
: IMMEDIATE SOURCE:
: CLONE: Fig.5a (S. faecalis)
: US-08-737-716-13

Query Match          5.6%; Score 106; DB 2; Length 671;
Best Local Similarity 23.6%; Pred. No. 0.046;
Matches 73; Conservative 39; Mismatches 103; Indels 94; Gaps 17;

QY 41 RNESSPKNLT-LNNHITAYSESHLYDRMTFLCLSSNHTLNGACPTSENPSSSV----- 94
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 309 RYATDPSTYNAKLNNVITAY--NLTOYDTPSSGCGTGNPGTGSNNQSGTNTYTYVK 366
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 95 SGET--NTTLOF---TKRSL--IKRELQIKGYKQLFKSVNCPSSGLTNSAHENCNKN 146
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 367 SGLTLTKIAAOYGVSVANMLRSMNGISGLIFVGOKLIVKGA--SGNTGGSGNGSGNN 423
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 147 AASGASLYLYPAGELKMLPGGIWDATLKLVRKRYSEYTGTYTINTIKLTDKGNIOI 206
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 424 -QSGTNTYTYTKSGDTLN-----KINAOYG-----VTV-----ANLRS 455
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 207 WLPOKRSARVDNL-----RPTGGGYIGRNSVDMCFDYGTSSSSLEIRFODNN 258
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 456 W-----NGISGLIFVGOKLIVKKGTSNT-----GSSNGGS-----NNN 491
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 259 PKSDGKFLRKINDTKELA--YTLS-----LLAGSLRPTNTGSLINDA 303
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 492 OSGTNTYTYTISGDTLNIAOYGVSVANLRSWNGISGLIFAGOKIIVKKGTSNGTGS 551
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 304 ASLETNNMR 312
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 552 SNGSGSNNNO 560
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 13
: -09-206-942-35
: Sequence 35, Application US/09206942
: Patent No. 6432669
: GENERAL INFORMATION:
: APPLICANT: Loosmore, Sheena M.
: APPLICANT: Yang, Yan-ping
: APPLICANT: Klein, Michel H.
: TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High
: TITLE OF INVENTION: Molecular Weight Proteins
: FILE REFERENCE: 1038-861 MTS:jb
: CURRENT APPLICATION NUMBER: US/09/206,942
: CURRENT FILING DATE: 1998-12-08
: EARLIER APPLICATION NUMBER: 09/167,568
: EARLIER FILING DATE: 1998-10-07
: NUMBER OF SEQ ID NOS: 95
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 35
: LENGTH: 915
: TYPE: PRT
: ORGANISM: Haemophilus Influenzae
: US-09-206-942-35

Query Match          5.4%; Score 101; DB 4; Length 915;
Best Local Similarity 21.1%; Pred. No. 0.25;
Matches 84; Conservative 48; Mismatches 129; Indels 138; Gaps 19;
```

```

QY 21 VSADKIPGDESTINIEGPR-----DRNESSPKH-----NILN-----NHITAYS 59
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 252 ISSDKVNITERITIKAGVNGDSNDEATSANLTKTEKLTNDLNSGFNKAEITAKD 311
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 60 ESH-FLYD-----RMTF-----LCSSHN--TLNGACPTSENPSSSVSGEFNI 100
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 312 NSNLTIGNSDAGNTDAKRVFSNVKDSKISASDHNVTLNSKVEFSGDTDEGCGNNNT 371
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 101 TLQFTKRSLLIKRELQIKGYKOL--LFKSYNCPSSLTFLNSAHFNCKNNAASGASLYLYI 157
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 372 GLTITAKKVVTVNN--NITSHKTVNITASENVTTKAGTTIN-----ATTGSVEVTA 419
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 158 PAGELKNLPFGGIMDATLKLVRKRYSEYTGTYTINTIKLTDKGNIOIWLPOKRSARV 217
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 420 KTGDIK-----GGI-----EBSNGVNI----- 437
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 218 DLNLRPTGGGYI-----GRNSVDMCFYDGYSTNSSLEIRFODNNPKSDGKFLRKIN 271
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 438 -----TASGDTLNSNITGQN-VTVAAASGAVTTTKGSTINATGNANITTK--TGEIN 488
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 272 DTKELIAYTLLSLAGKSLPTNGTSLNIADAASLETNNRITAVTMEISVPVLCWPCR 331
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 489 GEVKSASGNVNITASGNTLNSNITGQVTVTVAN-----SGAITTEGST----- 533
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 332 LQLDKAVENPEAGQYMGNNIN-----VFTFPSSQTL 361
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 534 --INATGDNATTOTGTGNGINKVESSSGSVTLIATGOTL 570
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 14
: US-09-206-942-37
: Sequence 37, Application US/09206942
: Patent No. 6432669
: GENERAL INFORMATION:
: APPLICANT: Loosmore, Sheena M.
: APPLICANT: Yang, Yan-ping
: APPLICANT: Klein, Michel H.
: TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High
: TITLE OF INVENTION: Molecular Weight Proteins
: FILE REFERENCE: 1038-861 MTS:jb
: CURRENT APPLICATION NUMBER: US/09/206,942
: CURRENT FILING DATE: 1998-12-08
: EARLIER APPLICATION NUMBER: 09/167,568
: EARLIER FILING DATE: 1998-10-07
: NUMBER OF SEQ ID NOS: 95
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 37
: LENGTH: 1222
: TYPE: PRT
: ORGANISM: Haemophilus influenzae
: US-09-206-942-37

Query Match          5.4%; Score 101; DB 4; Length 1222;
Best Local Similarity 21.1%; Pred. No. 0.4;
Matches 84; Conservative 48; Mismatches 129; Indels 138; Gaps 19;
```

```

QY 218 DLNLRPGGGYI-----GNSVDMCFDYGSINSSLSLEIRFDNNPKKDKGRFLRKIN 271
Db 745 -----TASGPTLWNSITGQN-VTVAAASGAVTTTGGSTINATGTGANITTK--TGEIN 795
QY 272 DDTKEIATLTLLAGKSLPTNGTSLINADAALETNNRITPTVTPETISVPLCLWPGR 331
Db 796 GEVKSASGNVITASGNLTNLSNITGQNVYTAN-----SGAITTEGST----- 840
QY 332 LQLDAKVENPEAGQYGMNIN-----VTFPSSQTL 361
Db 841 --INATITGDANITQTGTGINKVSVSSSGSYLLATGQTL 877

```

RESULT 15  
US-09-206-942-34

; Sequence 34, Application US/09206942

; Patent No. 6432669

; GENERAL INFORMATION:

APPLICANT: Loosmore, Sheena M.

APPLICANT: Yang, Yan-Ping

APPLICANT: Klein, Michel H.

TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High  
; TITLE OF INVENTION: Molecular Weight Proteins

; TITLE OF INVENTION: Molecular Weight Proteins

CURRENT APPLICATION NUMBER: US/09/206,942

CURRENT FILING DATE: 1998-12-08

EARLIER APPLICATION NUMBER: 09/167,568

EARLIER FILING DATE: 1998-10-07

; NUMBER OF SEQ ID NOS: 95

```

; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 34

```

; SEQ ID NO 34  
LENGTH: 12

LENGTH: 1228  
TYPE: PPT

ORGANISM:

ORGANISM: Haemophilus influenzae  
ITS-09-206-942-34

US-09-206-942-34

```
Query Match      5.48; Score 101; DB 4; Length 1228;
```

Matches 84; Conservative 48; Mismatches 129; Indels 138; Gaps 19;

[illegible]

Search completed: December 6, 2002, 18:18:30  
Job time : 22 secs



DR WPI; 2002-049280/06.  
DR N-PSDB; AAI70763, AAI70780.  
XX  
PT New nucleotide sequence, useful as immunogenic agent for generating  
PT immune response against recombinant product of the operon, comprises  
PT csa operon which encodes enterotoxigenic Escherichia coli-CS4 pili -  
XX  
PS Claim 10; Page 59; 81pp; English.  
XX  
CC The present sequence is that of the tip associated protein CsaE  
CC of enterotoxigenic Escherichia coli (ETEC) strain E11881A. CsaE is  
CC encoded by the csa gene (see AAI70763) of the E. coli E11881A csa  
CC operon. This operon includes 5 contiguous genes, csaA-csaE, which  
CC encode the synthetase of ETEC-CS4 pili. It has been expressed in  
CC attenuated Shigella strain CVD1204 guaba, constructing the Shigella  
CC expressing CS4 fimbriae vaccine strain CVD1204 (PGA2-CS4). The  
CC CsaE protein has a calculated mol.wt. of 40102.4 and a theoretical  
CC pI of 8.74. It shows homology to similar proteins from other ETEC  
CC fimbriae, recombinant CsaA-CsaE polypeptides are used in claimed  
CC immunogenic compositions to generate an immune response in a  
CC subject. These prevent ETEC colonisation, and hence protect  
CC against diarrhoea.  
XX  
XX Sequence 361 AA:  
Query Match 100.0%; Score 1886; DB 23; Length 361;  
Best Local Similarity 100.0%; Pred. No. 4e-169;  
Matches 361; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MNKILFIETLFESSVLTFFAVSADKIPGDESITNIFGPRDRN--ESSPKHNLNNHTAY 60  
DB 1 MNKILFIETLFESSVLTFFAVSADKIPGDESITNIFGPRDRN--ESSPKHNLNNHTAY 60  
QY 61 SHLYDRMTFLCASHNTLNGACPTSENPSSSVSGETNITLQFTKRSLLKRELQIKGY 120  
DB 61 SHLYDRMTFLCASHNTLNGACPTSENPSSSVSGETNITLQFTKRSLLKRELQIKGY 120  
QY 121 KOLLFKSVNCPGSLTNSAHFNCKNAASGASLYLTPAGELKLPFGGIMDAFLKLRVK 180  
DB 121 KOLLFKSVNCPGSLTNSAHFNCKNAASGASLYLTPAGELKLPFGGIMDAFLKLRVK 180  
QY 181 RRYSETGTITNTIKLTQKGNIOIWLPOKSPARVDLNLPTGGGTYIGRNSVDKCFY 240  
DB 181 RRYSETGTITNTIKLTQKGNIOIWLPOKSPARVDLNLPTGGGTYIGRNSVDKCFY 240  
QY 241 DGYSTNSSLEIRFODNNPKSDGKFLRKINDTKELAYTLSELLAKSLPTNGSLNI 300  
DB 241 DGYSTNSSLEIRFODNNPKSDGKFLRKINDTKELAYTLSELLAKSLPTNGSLNI 300  
QY 301 ADAASLETNNRITAVTMPETISVPLCPGRLQDLDAKVENPEAGQYMGNTINVTTPSSQT 360  
DB 301 ADAASLETNNRITAVTMPETISVPLCPGRLQDLDAKVENPEAGQYMGNTINVTTPSSQT 360  
QY 361 L 361  
DB 361 L 361  
RESULT 2  
ID AAY22326 standard; Protein: 364 AA.  
XX  
XX AAY22326;  
XX  
XX 22-SEP-1999 (first entry)  
XX  
XX Piliin protein Csd.  
XX  
XX CS2 gene cluster; Cota; Cota; Cota; Cota; Piliin protein; immunogen;  
XX enterotoxigenic E. coli; human upper intestine; diarrhoeal disease;  
XX  
XX Escherichia coli.

XX  
XX US5932715-A.  
XX  
XX 03-AUG-1999.  
XX  
XX 07-JUN-1995; 9505-0483101.  
XX  
XX 07-JUN-1995; 9505-0483101.  
XX  
XX (UYEM-) UNIV EMORY.  
XX  
XX Caron J, Froehlich B, Scott JR.  
XX  
XX WPI; 1999-443623/37.  
XX  
XX N-PSDB; AAX84848.  
XX  
PT Isolated nucleic acids encoding Escherichia coli CS2 piliin proteins  
PT useful for vaccinating against diarrhoeal diseases caused by  
PT Escherichia coli  
XX  
XX Claim 3; Column 45-48; 35pp; English.  
XX  
CC This sequence represents a CS2 piliin protein of the invention, encoded by  
CC (which also encodes Cota, Cota, and Cota). CS2 piliins are long  
CC proteinaceous molecules thought to mediate attachment of enterotoxigenic  
CC E. coli (ETEC) to and/or promote colonisation of the human upper  
CC intestine. The CS2 gene cluster may be used to produce immunogens for  
CC vaccinating patients against diarrhoeal diseases caused by ETEC bacteria.  
CC This type of enteric infection is a major cause of death among infants in  
CC developing countries and in immunocompromised (e.g. Acquired Immune  
CC Deficiency Syndrome (AIDS)) or elderly adults. The vaccine comprises more  
CC than 1 antigenic determinant (epitopes) from more than 1 piliin type to be  
CC effective against more than 1 type of ETEC infection.  
XX  
XX Sequence 364 AA:  
Query Match 49.2%; Score 927.5; DB 20; Length 364;  
Best Local Similarity 48.1%; Pred. No. 9.1e-79;  
Matches 176; Conservative 65; Mismatches 118; Indels 7; Gaps 3;  
QY 1 MNKILFIETLFESSVLTFFAVSADKIPGDESITNIFGPRDRN--ESSPKHNLNNHTAY 58  
DB 1 LKVIIFVLSMFLCSQVYGGSHNTVEAGSINKTISGPIRDSAAASPAHIFHEHAGY 60  
QY 59 SESHTLYDRMTFLCASHNTLNGACPTSENPSSSVSGETNITLQFTKRSLLKRELQIK 118  
DB 61 NKDHSLEDRMTFLCASHNTLNGACPTSENPSSSVSGETNITLQFTKRSLLKRELQIK 118  
QY 119 GYKQLEKSVNCPGSLTNSAHFNCKNAASGASLYLTPAGELKLPFGGIMDAFLKLRVK 175  
DB 119 GYKQLEKSVNCPGSLTNSAHFNCKNAASGASLYLTPAGELKLPFGGIMDAFLKLRVK 175  
QY 176 KLRVRRRYSETGTITNTIKLTQKGNIOIWLPOKSPARVDLNLPTGGGTYIGRNSV 235  
DB 176 KLRVRRRYSETGTITNTIKLTQKGNIOIWLPOKSPARVDLNLPTGGGTYIGRNSV 235  
QY 239 EMCYDGYSTNSSLEIRFODNNPKSDGKFLRKINDTKELAYTLSELLAKSLPTNGSLNI 298  
DB 239 EMCYDGYSTNSSLEIRFODNNPKSDGKFLRKINDTKELAYTLSELLAKSLPTNGSLNI 298  
QY 296 TSLNIAADAASLETNNRITAVTMPETISVPLCPGRLQDLDAKVENPEAGQYMGNTINVTTP 355  
DB 296 TSLNIAADAASLETNNRITAVTMPETISVPLCPGRLQDLDAKVENPEAGQYMGNTINVTTP 355  
QY 356 PSSQTL 361  
DB 356 PSSQTL 361  
RESULT 3  
ID AAB45919 standard; Protein: 359 AA.  
XX  
XX

AC AAB45919;  
XX  
XX 23-MAR-2001 (first entry)  
XX  
DE S. enterica serovar Typhi tcfD fimbrial subunit protein.  
XX  
XX Fimbrial protein; saf; tcf; vaccine; gene therapy; immunization;  
KW tcf insert; detection.  
XX  
OS Salmonella typhi.  
XX  
PN MO200073336-A1.  
XX  
PD 07-DEC-2000.  
XX  
XX 26-MAY-2000; 2000WO-SE01079.  
XX  
XX 28-MAY-1999; 99SE-0001961.  
XX  
XX (ACTI-) ACTIVE BIOTEC AB.  
XX  
XX Folkeesson A, Normark S, Lofdah S;  
XX WPI: 2001-061512/07.  
DR N-PSDB; AAC82926.  
XX  
XX Fimbriae proteins of Salmonella enterica subspecies I bacteria, useful  
PT for producing vaccines against the bacterial subspecies and for  
XX detecting the bacteria  
XX  
XX Disclosure; Page 72-73; 77pp; English.  
XX  
XX This invention describes the novel proteins (saf and tcf) (1) encoded by  
CC a DNA sequence of a gene encoding the precursor of the saf fimbriae unit  
CC of Salmonella enterica subspecies I or a DNA sequence of the gene  
CC encoding the tcf fimbriae unit of S. enterica subspecies I serovar Typhi  
CC The products of the invention can be used as vaccines or for gene  
CC therapy. Such vaccines are useful for protection against diseases caused  
CC by S. enterica subspecies I or S. enterica subspecies I serovar Typhi.  
CC The saf and tcf proteins from S. enterica subspecies I bacteria are  
CC useful for active or passive immunization in mammals. The nucleotide  
CC sequences are useful for constructing vectors for use as vaccines for  
CC insertion into attenuated bacteria in constructing a recombinant viral  
CC vaccine, or for direct inoculation of a nucleic acid vaccine. The protein  
CC or antigenic fragments, nucleic acid sequences, and antibodies are useful  
CC in molecular diagnostic assays for the detection of S. enterica  
CC subspecies I.  
XX  
XX Sequence 359 AA;  
XX  
XX Query Match 14.6%; Score 275.5; DB 22; Length 359;  
XX Best Local Similarity 29.7%; Pred. No. 2.6e-17;  
XX Matches 90; Conservative 48; Mismatches 130; Indels 35; Gaps 11;

OY 358 SQT 360  
DB 356 TQT 358  
RESULT 4  
ID AAY50814 standard; Protein: 770 AA.  
XX  
XX AAY50814;  
XX  
XX 17-FEB-2000 (first entry)  
XX  
XX D. melanogaster acetyl-choline receptor protein from clone Da7.  
XX  
XX Acetyl-choline receptor; nicotinic; insect; insecticide; screening;  
KW neurotransmission; plant protection agent; conductance; AChR.  
XX  
XX Drosophila melanogaster.  
XX  
XX DE19819829-A1.  
XX  
XX 11-NOV-1999.  
XX  
XX 04-MAY-1998; 98DE-1019829.  
XX  
XX 04-MAY-1998; 98DE-1019829.  
XX  
XX 04-MAY-1998; 98DE-1019829.  
XX  
XX (FARB ) BAYER AG.  
XX  
XX Adamczewski M, Oellers N, Schulte T;  
XX WPI: 2000-014207/02.  
DR N-PSDB; AA224475.  
XX  
XX New nucleic acid encoding a nicotinic acetylcholine receptor from  
PT insects, used to identify potential insecticides  
XX  
XX Example 1a: Page 12-14; 26pp; German.  
XX  
XX This invention describes a novel nucleic acid (NA) encoding a nicotinic  
CC acetyl-choline receptor (1) from insects which can be used as an  
CC insecticide. Inhibitors of (1) interfere with neurotransmission. (1)  
CC (also vectors containing it, its regulatory regions, and antibodies  
CC directed against (1)-encoded proteins) are used to screen for: (a) plant  
CC protection agents that alter conductance of AChR, potentially useful as  
CC insecticides, or (b) genes which encode polypeptides that are involved in  
CC formation of functionally related AChR in insects. (1) are also used to  
CC isolate and characterize the specified regulatory regions and for  
CC recombinant production of (1). This sequence represents an  
CC acetyl-choline receptor isolated from Drosophila melanogaster.  
XX  
XX Sequence 770 AA;  
XX  
XX Query Match 6.3%; Score 119.5; DB 21; Length 770;  
XX Best Local Similarity 20.4%; Pred. No. 0.041;  
XX Matches 83; Conservative 48; Mismatches 140; Indels 135; Gaps 20;

Db 396 PVLVLM-----YVSADEGPDCTQTNNVVR--NNGSC-LYVPEIGKSTCKIDITTFP---- 444

QY 227 GYVIGRNSVDKMF---YDGYSTNSSLEIRFQDNNPKS-----DGKFLRKINDTKE 276

Db 445 ---FDDQREKMFSGWTVYGF-----QDLQLODERGSGISSVVLNGEMLLGVPEKRRNE 496

QY 277 IAYT-----LGLLAGKSLPTNGTSLNIADA 303

Db 497 IYVNCCEPEYIDITFAIIRRLTYFFENLIIPCVLIASMALGFTLPPDGEKLSLGYT 556

QY 304 ASLE-TNNMRITAVTMEIS--VPV-----LCW-----PGR 331

Db 557 ILLSLTVFLNMVAETMPATSDAIPMIRIVFLCWLPEWILRMSRPR 602

RESULT 5  
AAR47575

ID AAR47575 standard; Protein; 650 AA.

XX AAR47575;

DT 19-JUL-1994 (first entry)

XX

DE Alpha-agglutinin of *Saccharomyces cerevisiae*.

XX

KM Immobilisation; enzyme; cell wall; alpha agglutinin; AGA 1; FLO 1;

KW Major cell wall protein; glycosyl-phosphatidyl-inositol;

KM anchoring protein; alpha factor; alpha-agglutinin; invertase;

KW inulinase; alpha-amylase; *Saccharomyces cerevisiae*; catalysis.

KM enzymatic process; fermentation; biodegradation; catalysis.

XX

OS *Saccharomyces cerevisiae*.

XX

PN MO9401567-A.

PD 20-JAN-1994.

XX

PF 07-JUL-1993; 93WO-EP01763.

XX

PR 08-JUL-1992; 92EP-0202080.

PR 14-DEC-1992; 92EP-0203899.

XX

PA (UNITL ) UNILEVER NV.

PA (UNITL ) UNILEVER PLC.

XX

PI K11s FM, Schreuder MP, Toschka H, Verrips CT;

XX

WP: 1994-035071/04.

N-PSDB: AA054012.

XX

PT Immobilisation of enzymes to microbial cell wall - by prodn. of

PT fusion protein of enzyme linked to anchoring protein

XX

PS Example 1; Page 32-39; 99pp; English.

XX

CC The alpha-agglutinin is used in a method to immobilise enzymes to a

CC microbial cell wall. The coding sequence is used in the production

CC of a recombinant polynucleotide which comprises a structural gene

CC encoding a protein with catalytic activity and at least part of a

CC gene encoding at least the C-terminus of a protein capable of

CC anchoring in a eukaryotic or prokaryotic cell wall. The anchoring

CC fragment of protein is selected from alpha agglutinin, AGA 1, FLO 1,

CC major cell wall protein of lower eukaryotes or a proteinase of

CC lactic acid bacteria. The recombinant polynucleotide preferably

CC also comprises a sequence encoding a signal peptide to ensure

CC secretion of the expressed product. The signal peptide is

CC preferably derived from glycosyl-phosphatidyl-inositol, anchoring

CC protein, alpha factor, alpha-agglutinin, invertase or inulinase,

CC alpha-amylase of *Bacillus* or proteinases of lactic acid bacteria.

CC The host microorganism can be used for performing enzymatic

CC processes on an industrial scale.

XX

XX Sequence 650 AA:

XX

PT

Query Match 6.0%; Score 114; DB 15; Length 650;

Best Local Similarity 20.9%; Pred. No. 0.1;

Matches 98; Conservative 52; Mismatches 147; Indels 172; Gaps 23;

QY 4 ILFFITLFFSSVVL---FFPA-----VSADKTPG-----DESTINFGPPDRNE--- 43

Db 8 ILMFSLALASAININDIFFSNIETPLTANKOPDQGWATPDEFSLADASSIREGDEFTL 67

QY 44 SSP---KHNILNNHTA-----YSESHTYDRMTFLC-----LSSHNTLN 80

Db 68 SMPHYTRIKLNNSSQTAATISLADGTEAFKCYVSOQAAYILENTTFTCTAONDLSNTITD 127

QY 81 GACPTEENPSSSSVSGETNITLQTEKRSLIKRELQIKGYKOLLFSVNC-PSGLTLNSA 139

Db 128 GSITFSLNFDGSSVEYEL-----ENAKPEKSGPMLVTLGNQMSDVNFPDPAAPTEVNF 182

QY 140 HFNCNKNAAASGASLYL--YIPAGELKNLPGGIWDATLKLRYKRRYSFYGYTTINIITK 197

Db 183 HSGRSTGYGSFESYHLMGMCPNGVF---LGG---TERI-----DYDSNNNVLD- 225

QY 198 LTKDKNIOI-----WLPQEKSDARVDLNRPTGGGYIGRN---SVDMGFYDGC--- 242

Db 226 --DCSSVQVYSSNDFNDMPFPOSYNPTNADV-----TCFESNIMTLDEKLDGEML 275

QY 243 -----YSTNSSLE--IRFQDNN--PKSDGKF 265

Db 276 WYNALQSLPANVTIDHALFQYCTCLDIANTTYAQFTTREFIYQGRNLGTASAKSS 335

QY 266 YLRKINDTKET--AYTIS-----LLLAGKSLPTNGTSLNIADA 304

Db 336 FISTTTTDLTSLNTSAYSGSISYVETGNRTSEVISHVVTSTKLSPTATSLTIAQTS 395

QY 305 SLETMNN-----RLTAVTMEISVPVLQWPGRL 332

Db 396 IYSTDSNITVGDIDHTTSEVISDVEITISRETAStVAAPTSTTGWTGAM 444

RESULT 6  
AAB01833

ID AAB01833 standard; Protein; 1005 AA.

XX

AC AAB01833;

XX

DT 11-SEP-2000 (first entry)

XX

DE Haemophilus influenzae strain K21 mature HmW2A protein, SEQ ID NO:41.

XX

KW Mature HmW protein; hmw gene; hmwA1; hmwA2; high molecular weight;

KW non-typeable Haemophilus influenzae; NTN1; non-encapsulated;

KW recombinant production; Escherichia coli; antibacterial; vaccine;

KW human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;

XX

XX detection; diagnosis.

XX

OS Haemophilus influenzae strain K21.

XX

PN WO200020609-A2.

XX

PD 13-APR-2000.

XX

PF 07-OCT-1999; 99WO-CA00938.

XX

PR 07-OCT-1998; 98US-0167568.

PR 08-DEC-1998; 98US-0206942.

XX

PA (CONN-) CONNAUGHT LAB LTD.

XX

PI Loosemore SM, Yang Y, Klein MH;

XX

DR WP: 2000-303789/26.

DR N-PSDB: AAA52182.

XX

PT Nucleic acid molecule for producing recombinant high molecular weight



PT proteins of Haemophilus which are used as a vaccine to provide  
 PT protection against Haemophilus induced diseases in humans -  
 PS Claim 8; Fig 21A-O; 307pp; English.

CC The invention relates to the recombinant production of Haemophilus  
 CC influenzae high molecular weight (HMW) proteins in Escherichia coli. The  
 CC expression construct used to effect recombinant expression comprises a  
 CC promoter functional in E. coli (e.g., the 7' promoter) operably linked  
 CC to a modified hmwaBC operon from a non-typeable (non-encapsulated) H.  
 CC influenzae (NTH1). Most HMW-expressing NTH1 strains contain two hmw gene  
 CC clusters termed hmwa1BC and hmwa2ABC. Each hmwaBC operon comprises hmwa,  
 CC hmwb and hmwc genes. The hmwa genes encode the structural HMWA proteins  
 CC and the hmwb and hmwc genes encode accessory proteins which are  
 CC responsible for post-translational processing and secretion of the HMWA  
 CC proteins. The modified hmwaBC operon used in the expression construct of  
 CC the invention contains an A gene modified such that it encodes only the  
 CC mature HMWA. The invention also discloses hmwa genes (AA52175-452198)  
 CC and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae  
 CC strains J09c, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and  
 CC vectors are used for the production of recombinant H. influenzae HMW  
 CC proteins which can be used as vaccines to mediate a humoral or  
 CC cell-mediated immune response to provide protection against diseases in  
 CC humans caused by H. influenzae (e.g., otitis media, epiglottitis,  
 CC pneumonia and tracheobronchitis). The HMW proteins are also useful as  
 CC antigens in immunoassays for detecting antibodies against Haemophilus,  
 CC HMW proteins and/or HMW peptides. The nucleotide sequences encoding the  
 CC HMW proteins can be used to isolate and clone hmw genes from other  
 CC non-typeable strains of Haemophilus via hybridisation reactions. The  
 CC present sequence represents a mature HMWA protein from a non-typeable  
 CC strain of H. influenzae.

CC Sequence 1005 AA;

Query Match 5.9%; Score 111.5; DB 21; Length 1005;  
 Best Local Similarity 20.3%; Pred. NO. 0.34;

Matches 68; Conservative 65; Mismatches 147; Indels 55; Gaps 14;

QY 60 ESHHTLYDRMTFLCLSSHNTLNACPTSENPSSSVSGENITLOTFEKS-LIKRELOIK 118  
 DB 363 EKNAIFSTHNLITLGGVNTLGGENSSNITKGININISKANVTLOAHAGTSHLDKKE--- 418  
 QY 119 GYKOLLFESVNCPSGLTL--NSAHFNCNKNAAGSLVLYIPAGEL-KNLPGGIW--DA 173  
 DB 419 --RTLLTGVNSVGNLNTIGSNAHIDGNLSIAESAKF-----QGTNNNLNTGFTTNG 471  
 QY 174 TLKLVKRRYSEYGTNTINIKLTD-----KGNL-----QIWLPOFKSDARVD 218  
 b 472 TADINIKGVNVVLOGDITNNGMLNTTNASVNOQTIIINGNITNKKGDLINIKDKANAEIO 531  
 QY 219 L--NLRPNGCGTYIGRNSVDMCFYDGYSTNSSSLEIRPDNNPKSP-GKFIYLRKINDTRK 275  
 DB 532 ICGNISQKRGKNTLTISSDKINT-----TKRIETKADTDGNSDSGVASNAANLITKTK 582  
 QY 276 ELAYTLISLLLAG--KSLTPNGTSINIAADAASLETNMRITAVTPEISVPLCPGRL 332  
 DB 583 ELTLFDNLNISEGFNAETAKDNDLIGKASSDMSNAQIIFEDVKOSKIS--AGNHNV 640  
 QY 333 QIDAKVENPEA-----GOYKGNINVTFTPSSQTL 361  
 DB 641 TLNSKVEITSNSDGTGNGSDNNIGLITISAKDVTY 675

RESULT 7

AA01832 standard; Protein; 1011 AA.

AA01832;  
 11-SEP-2000 (first entry)  
 Haemophilus influenzae strain K21 HMW2A protein, SEQ ID NO:39.

KW HMW protein; hmw gene; hmwa1; hmwa2; high molecular weight;  
 KW non-typeable Haemophilus influenzae; NTH1; non-encapsulated;  
 KW recombinant production; Escherichia coli; antibacterial; vaccine;  
 KW human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;  
 KW detection; diagnosis.

Haemophilus influenzae strain K21.

WO200020609-A2.

13-APR-2000.

07-OCT-1999; 99WO-CA00938.

07-OCT-1998; 98US-0167568.

08-DEC-1998; 98US-0206942.

(CONN-) CONNAUGHT LAB LTD.

Loosemore SM, Yang Y, Klein MH;

WPL; 2000-303789/26.

N-PSDB; AAA52181.

Nucleic acid molecule for producing recombinant high molecular weight  
 PT proteins of Haemophilus which are used as a vaccine to provide  
 PT protection against Haemophilus induced diseases in humans -

Claim 12; Fig 21A-O; 307pp; English.

CC The invention relates to the recombinant production of Haemophilus  
 CC influenzae high molecular weight (HMW) proteins in Escherichia coli. The  
 CC expression construct used to effect recombinant expression comprises a  
 CC promoter functional in E. coli (e.g., the 7' promoter) operably linked  
 CC to a modified hmwaBC operon from a non-typeable (non-encapsulated) H.  
 CC influenzae (NTH1). Most HMW-expressing NTH1 strains contain two hmw gene  
 CC clusters termed hmwa1BC and hmwa2ABC. Each hmwaBC operon comprises hmwa,  
 CC hmwb and hmwc genes. The hmwa genes encode the structural HMWA proteins  
 CC and the hmwb and hmwc genes encode accessory proteins which are  
 CC responsible for post-translational processing and secretion of the HMWA  
 CC proteins. The modified hmwaBC operon used in the expression construct of  
 CC the invention contains an A gene modified such that it encodes only the  
 CC mature HMWA. The invention also discloses hmwa genes (AA52175-452198)  
 CC and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae  
 CC strains J09c, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and  
 CC vectors are used for the production of recombinant H. influenzae HMW  
 CC proteins which can be used as vaccines to mediate a humoral or  
 CC cell-mediated immune response to provide protection against diseases in  
 CC humans caused by H. influenzae (e.g., otitis media, epiglottitis,  
 CC pneumonia and tracheobronchitis). The HMW proteins are also useful as  
 CC antigens in immunoassays for detecting antibodies against Haemophilus,  
 CC HMW proteins and/or HMW peptides. The nucleotide sequences encoding the  
 CC HMW proteins can be used to isolate and clone hmw genes from other  
 CC non-typeable strains of Haemophilus via hybridisation reactions. The  
 CC present sequence represents an HMWA protein from a non-typeable strain of  
 CC H. influenzae.

Sequence 1011 AA;

Query Match 5.9%; Score 111.5; DB 21; Length 1011;  
 Best Local Similarity 20.3%; Pred. NO. 0.35;

Matches 68; Conservative 65; Mismatches 147; Indels 55; Gaps 14;

QY 60 ESHHTLYDRMTFLCLSSHNTLNACPTSENPSSSVSGENITLOTFEKS-LIKRELOIK 118  
 DB 363 EKNAIFSTHNLITLGGVNTLGGENSSNITKGININISKANVTLOAHAGTSHLDKKE--- 424  
 QY 119 GYKOLLFESVNCPSGLTL--NSAHFNCNKNAAGSLVLYIPAGEL-KNLPGGIW--DA 173  
 DB 425 --RTLLTGVNSVGNLNTIGSNAHIDGNLSIAESAKF-----QGTNNNLNTGFTTNG 477  
 QY 174 TLKLVKRRYSEYGTNTINIKLTD-----KGNL-----QIWLPOFKSDARVD 218

```

Db 478 TADINIKQGVKLOGDITNNGNINITTNASVNOKTIINGNITNKGDNLNIKAMAEIQ 537
QY 219 L--NLRPFGGTYIGRNSVDMCFYDGYSTNSSLEIRFQDNPKSD--GKFYLKRIKINDTK 275
Db 538 IGNISQKEGNLTISDKIYI-----TKRIEKADTDGNSDGSVANSANLTIKTK 588
QY 276 ELAVYLSLLAG--KSLPTNGTSLNIADAASLETNNMNRITAVTPEISVPLYCWPGR 332
Db 589 ELTLTDNLNINSGFNKAIEITAKDNSDLIGKASSDNSNAKQITFDKVKSKIS--AGNHNV 646
QY 333 QLDKAYENPEA-----GQYMGNIINTFTPSQTL 361
Db 647 TLNSKVETSNDSGTGSDDNNGILGISAKDVTY 681

```

## RESULT 8

AA041732  
ID AAR41732 standard; Protein; 1529 AA.

XX AAR41732;

XX 26-APR-1994 (first entry)

XX High molecular weight protein 4 (HMW4).

XX HMW: high molecular weight protein; virus; vaccine; influenza;  
epitope; immunity; haemophilus influenzae.

XX Haemophilus influenzae.

XX WO9319090-A.

XX 30-SEP-1993.

XX 16-MAR-1993; 93WO-US02166.

XX 16-MAR-1992; 92GB-0005704.

XX (BARE/) BARENKAMP S J.  
(INRM ) INSERM INST NAT SANTE & RECH MEDICALE.

XX Barenkamp SJ;

XX WPI; 1993-320683/40.

XX N-PSDB; AA049511.

XX High molecular weight surface proteins - of non-typeable  
haemophilus which exhibit immunogenic properties

XX Claim 6; Figure 10; 100pp; English.

XX The isolation and purification of the high molecular weight protein  
CC enables the identification of the major protective epitopes of the  
CC protein by conventional epitope mapping. These epitopes can then be  
CC synthesised using standard techniques and incorporated into fully  
CC synthetic or recombinant vaccines.

XX Sequence 1529 AA;

Query Match 5.9%; Score 110.5; DB 14; Length 1529;  
Best Local Similarity 20.7%; Pred. No. 0.8;  
Matches 72; Conservative 62; Mismatches 127; Indels 87; Gaps 18;

```

QY 48 HNLNHNHTAYSESHLYDWMFLCLSSHTLNLGACPTSENPS-----SVSGETNTTL 102
Db 744 YNEYSKH--AINSSHNL-----TILGAVNTLGG-----ENSSSITGNTINITNKAVTL 790
QY 103 QF-----TEKRSLIKRELQIKYKQLLFKSVNCPGSLTNSAHFNCKNNAAGASLY 154
Db 791 QADTSNNGKRTLTLLGNISVEGNLSLGCANANITGNLSI--AEDSTKGEAS----- 843
QY 155 LKIPAGELKLNLPFGTW--DATLKNRYKRYSTYGTYT-----INTIKLT----- 199

```

```

Db 844 -----DNLNITGTFTNNGTANINIKQGVKLOGDIIINKGGLNITNASTGTOKTLING 895
QY 200 ----DKGNIDIMLPQFSDARVDL--NLRPFGGTYIGRNSVDMCFYDGYSTNSSLEIR 253
Db 896 NITNEKGDINI--KNIKADEIQIGNISOKEGNLTISDKVNI-----TMDITIKAG 946
QY 254 FO-----DNNPKSDGKFLRKINDTKEIAYTSLLAG--KSLPTNGTSLNIADAASL 306
Db 947 VEGGRSDSSAEANANLTIQ-----TKELKLAGDLNINSGFNKAIEITAKNSDILITGNASG 1001
QY 307 ETMNRITAVTPEISVPLYCWPGR-RLQLDKAYENPEAGQYMGNIINT 353
Db 1002 MADAKK---VTFDKVKDKSISTDGHVNTLNSVEKTSNGSSNAGNDNST 1046

```

## RESULT 9

AA030292  
ID AAW30292 standard; Protein; 1601 AA.

XX AAW30292;

XX 14-APR-1998 (first entry)

XX Non-typeable Haemophilus high mol.wt. surface protein HMW4.

XX Non-typeable Haemophilus; high molecular weight surface protein;  
HMW4; immunogen; vaccine; otitis media.

XX Haemophilus influenzae strain 5.

XX WO936914-A1.

XX 09-OCT-1997.

XX 01-APR-1997; 97WO-US04707.

XX 01-APR-1996; 96US-0617697.

XX (BARE/) BARENKAMP S J.

XX Barenkamp SJ;

XX WPI; 1997-503038/46.

XX N-PSDB; AAT90993.

XX High molecular weight proteins of non-typeable Haemophilus  
influenzae - useful for vaccine production

XX Claim 1; Page 97-102; 183pp; English.

XX This protein comprises the high molecular weight surface protein  
CC HMW4 (123 kDa) of non-typeable Haemophilus influenzae strain 5 that  
CC has the immunological ability to protect against disease caused by  
CC a non-typeable Haemophilus strain and is characterised by at least  
CC one surface-exposed B-cell epitope that is recognised by monoclonal  
CC antibody A06. The HMW4 amino acid sequence was deduced from an  
CC isolated hmw4 gene (see AAT90993). HMW1 (see AAW30293), HMW2 (see  
CC AAW30294) and HMW3 (see AAW30291) have also been identified. A  
CC conjugate comprising HMW4 linked to an antigen, hapten or  
CC polysaccharide, and a synthetic peptide of 6-150 amino acids  
CC corresponding to at least protective epitope of HMW4 are also  
CC claimed. HMW proteins, conjugates and peptides can be used in  
CC vaccines, as immunogens for preparation of antibodies and as  
CC antigens for detection of these antibodies.

Sequence 1601 AA;



DR WPI: 2000-303789/26.  
 XX N-PSDB: AAA52184.  
 XX Nucleic acid molecule for producing recombinant high molecular weight  
 PT proteins of Haemophilus which are used as a vaccine to provide  
 PT protection against Haemophilus induced diseases in humans -  
 XX  
 PS Claim 8: Fig 22A-P: 307pp: English.  
 XX  
 CC The invention relates to the recombinant production of Haemophilus  
 CC influenzae high molecular weight (HMW) proteins in Escherichia coli. The  
 CC expression construct used to effect recombinant expression comprises a  
 CC promoter functional in E. coli (e.g., the T7 promoter) operably linked  
 CC to a modified hmwaBC operon from a non-typable (non-encapsulated) H.  
 CC influenzae (NTHI). Most HMW-expressing NTHI strains contain two hmw gene  
 CC clusters termed hmwa1ABC and hmwa2ABC. Each hmwaBC operon comprises hmwaA,  
 CC hmwaB and hmwaC genes. The hmwa genes encode the structural HMWA proteins  
 CC and the hmwaB and hmwaC genes encode accessory proteins which are  
 CC responsible for post-translational processing and secretion of the HMWA  
 CC proteins. The modified hmwaBC operon used in the expression construct of  
 CC the invention contains an A gene modified such that it encodes only the  
 CC mature HMWA. The invention also discloses hmwa genes (AAA52175-A52198)  
 CC and HMWA proteins (AAB01824-B01849) from the non-typable H. influenzae  
 CC strains Joyce, K1, K21, LCD2C, PMH1, 15 and 12. The nucleic acids and  
 CC vectors are used for the production of recombinant H. influenzae HMW  
 CC proteins which can be used as vaccines to mediate a humoral or  
 CC cell-mediated immune response to provide protection against diseases in  
 CC humans caused by H. influenzae (e.g., otitis media, epiglottitis,  
 CC pneumonia and tracheobronchitis). The HMW proteins are also useful as  
 CC antigens in immunassays for detecting antibodies against Haemophilus,  
 CC HMW proteins and/or HMW peptides. The nucleotide sequences encoding the  
 CC HMW proteins can be used to isolate and clone hmw genes from other  
 CC non-typable strains of Haemophilus via hybridisation reactions. The  
 CC present sequence represents a mature HMWA protein from a non-typable  
 CC strain of H. influenzae.  
 \*XX  
 SQ Sequence 1095 AA:  
 Query Match 5.7%; Score 107; DB 21; Length 1095;  
 Best Local Similarity 24.6%; Pred. No. 1;  
 Matches 71; Conservative 44; Mismatches 110; Indels 64; Gaps 15;  
 QY 45 SPKHNILNHHITAYSESHLYD-----RMFLCSSHNTLNGACPTSENPSSSS 93  
 DB 281 SIRPPIYSN---VHDGNHTLFNGNVSVLGGDVNFHFNASSSNHMHGVIKSONFNA5E 337  
 QY 94 VSGETNITLOFTEKRLIKRELQIKGYKOLLFKSVNCPGLTNSAIFNCKNKAASGASL 153  
 DB 338 GS-----SLRF-----KSEGSTRAF---TIESDLTLNATGNSLSLQVAGIDG 378  
 QY 154 YLYIPAGELKNLFPFGIWDATLKLVRKRRYSFYGYTTI---NITIKLTDKNGIOWLP 209  
 DB 379 NLQSKLIVANKNITFEF---GNITLAADKPIETIKGNTIVEGANVTLRSANYGNDKRAL- 434  
 QY 210 QEFSRDARVDLNAPTGGGTYIGRN-----SYDQCFYDGYSTNSSLFIRQDNNPKSD-- 262  
 DB 435 SIRGNVTNKGVLNVTGSAINIEKNLTVEGSAKFLANPNYSFNVSGL----FDQKGSNIS 490  
 QY 263 ---GKFLRKINDTKEIAYTL-----LLAGKSLPTNGSTLNAD 302  
 DB 491 IAKGGAHFKDIN-NITSLNITTTNSDSAYRTIIEG-NITNSNG-DLNIITD 536  
 RESULT 12  
 AAB01834  
 ID AAB01834 standard; Protein; 1101 AA.  
 XX  
 AC AAB01834;  
 XX  
 DT 11-SEP-2000 (first entry)  
 XX  
 DE Haemophilus influenzae strain LCD2C HMWA protein, SEQ ID NO:43.

KW HMW protein; hmw gene; hmwa1; hmwa2; high molecular weight;  
 KW non-typable Haemophilus influenzae; NTHI; non-encapsulated;  
 KW recombinant production; Escherichia coli; antibacterial; vaccine;  
 KW human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;  
 KW detection; diagnosis.  
 XX  
 OS Haemophilus influenzae strain LCD2C.  
 XX  
 PN WO200020609-A2.  
 XX  
 PD 13-APR-2000.  
 XX  
 PF 07-OCT-1999; 99WO-CA00938.  
 XX  
 PR 07-OCT-1998; 98US-0167568.  
 PR 08-DEC-1998; 98US-0206942.  
 XX  
 PA (CONN-) CONNAUGHT LAB LTD.  
 XX  
 PI Loosmore SM, Yang Y, Klein MH;  
 PI  
 DR WPI: 2000-303789/26.  
 DR N-PSDB: AAA52183.  
 XX  
 PT Nucleic acid molecule for producing recombinant high molecular weight  
 PT proteins of Haemophilus which are used as a vaccine to provide  
 PT protection against Haemophilus induced diseases in humans -  
 XX  
 PS Claim 12: Fig 22A-P: 307pp: English.  
 XX  
 CC The invention relates to the recombinant production of Haemophilus  
 CC influenzae high molecular weight (HMW) proteins in Escherichia coli. The  
 CC expression construct used to effect recombinant expression comprises a  
 CC promoter functional in E. coli (e.g., the T7 promoter) operably linked  
 CC to a modified hmwaBC operon from a non-typable (non-encapsulated) H.  
 CC influenzae (NTHI). Most HMW-expressing NTHI strains contain two hmw gene  
 CC clusters termed hmwa1ABC and hmwa2ABC. Each hmwaBC operon comprises hmwaA,  
 CC hmwaB and hmwaC genes. The hmwa genes encode the structural HMWA proteins  
 CC and the hmwaB and hmwaC genes encode accessory proteins which are  
 CC responsible for post-translational processing and secretion of the HMWA  
 CC proteins. The modified hmwaBC operon used in the expression construct of  
 CC the invention contains an A gene modified such that it encodes only the  
 CC mature HMWA. The invention also discloses hmwa genes (AAA52175-A52198)  
 CC and HMWA proteins (AAB01824-B01849) from the non-typable H. influenzae  
 CC strains Joyce, K1, K21, LCD2C, PMH1, 15 and 12. The nucleic acids and  
 CC vectors are used for the production of recombinant H. influenzae HMW  
 CC proteins which can be used as vaccines to mediate a humoral or  
 CC cell-mediated immune response to provide protection against diseases in  
 CC humans caused by H. influenzae (e.g., otitis media, epiglottitis,  
 CC pneumonia and tracheobronchitis). The HMW proteins are also useful as  
 CC antigens in immunassays for detecting antibodies against Haemophilus,  
 CC HMW proteins and/or HMW peptides. The nucleotide sequences encoding the  
 CC HMW proteins can be used to isolate and clone hmw genes from other  
 CC non-typable strains of Haemophilus via hybridisation reactions. The  
 CC present sequence represents an HMWA protein from a non-typable strain of  
 CC H. influenzae.  
 XX  
 SQ Sequence 1101 AA:  
 Query Match 5.7%; Score 107; DB 21; Length 1101;  
 Best Local Similarity 24.6%; Pred. No. 1;  
 Matches 71; Conservative 44; Mismatches 110; Indels 64; Gaps 15;  
 QY 45 SPKHNILNHHITAYSESHLYD-----RMFLCSSHNTLNGACPTSENPSSSS 93  
 DB 287 SIRPPIYSN---VHDGNHTLFNGNVSVLGGDVNFHFNASSSNHMHGVIKSONFNA5E 343  
 QY 94 VSGETNITLOFTEKRLIKRELQIKGYKOLLFKSVNCPGLTNSAIFNCKNKAASGASL 153  
 DB 344 GS-----SLRF-----KSEGSTRAF---TIESDLTLNATGNSLSLQVAGIDG 384  
 QY 154 YLYIPAGELKNLFPFGIWDATLKLVRKRRYSFYGYTTI---NITIKLTDKNGIOWLP 209

```

Db 385 NLQSLVANKNITFEG---GNITLADKKRIEIKNITVEKAGANVILRSANYGNDSAL- 440
OY 210 QPKSDARVDLNRPTGGGTGYIGRN-----SYDMCFYDGYSTNSSLEIRFQDNPKSD-- 262
Db 441 SLIGANTKNGNLVTGSAINIEKNLTVESGAKFLANPNISFNVSGL---FDQGRKSNIS 496
OY 263 ---GKFEYLRKINDDTKEIAYTLS-----LLAGKSLTFPTNGTSLNIAD 302
Db 497 IAKGGAHFPKNDIN-NTKSLNITNTSDSAYRTIIEG-NITNSNG-DLNIITD 542

RESULT 13
AAR85290
ID AAR85290 standard; Protein; 671 AA.
XX
AC AAR85290;
XX
DT 02-APR-1996 (first entry)
XX
DE Streptococcus faecalis autolysin.
XX
KW Lysin; autolysin; culture; lactic acid bacteria; fermentation;
KW cheese; foodstuffs; induction.
XX
OS Streptococcus faecalis.
XX
PN W09531561-A1.
XX
PD 23-NOV-1995.
XX
PF 12-MAY-1995; 95WO-NL00170.
XX
PR 12-MAY-1994; 94EP-0201353.
XX
PA (UNITL ) QUEST INT BV.
XX
PI Buist G, Kok J, Ledebor AM, Venema G;
XX
DR WPI; 1996-010946/01.
XX
PT Lysis of a culture of lactic acid bacteria in, e.g. cheese
PT production - by in situ prodn. of an auto:lysin, regulated by an
PT inducible promoter.
XX
PS Disclosure: Page 66-69; 103pp; English.
XX
CC In situ production of a homologous autolysin or a heterologous
CC autolysin from a food grade gram positive bacteria, can be used in
CC a process for the lysis of a culture of lactic acid bacteria. The
CC process can be used in the manufacture of products containing
CC cultures of lactic acid bacteria e.g. cheese, where the culture is
CC lysed following the completion of fermentation. The enhanced
CC induction of the autolysin is performed some hours after the
CC fermentation is finished. No extra lysis needs to be added and the
CC lysis does not need to be isolated or encapsulated. The time of
CC lysis can be precisely controlled. This is the Streptococcus
CC faecalis autolysin.
XX
SQ Sequence 671 AA;

Query Match 5.6%; Score 106; DB 17; Length 671;
Best local similarity 23.6%; Pred. No. 0.62;
Matches 73; Conservative 39; Mismatches 103; Indels 94; Gaps 17;
OY 41 RNSSSPKHHI-LNNHITAVSESHLTLYDRMTFLCLSHNTLNGACPTSENPSSSV----- 94
Db 309 RYATDPSYNAKLNKNVITAY--NLTYDTPSSGGTGNPPTGGGNNQSGTNYTYTVK 366
OY 95 SGEI--NITLQF-----TEKRSI--IKRELQIKYKOLLKFSVNCPSGTLNLSHFNCNNK 146
Db 367 SCGTLLKKAIAQGVSVANLRSMNGISGDLLEVGQKLIKKGA--SGNTGSGGSGGNNN 423
OY 147 AASGASLYIYPAGEIKNL.PFGIMDATLKLKRVKRYSETYGYTITITIKLTDKNIQI 206

```

```

Db 424 -QSGTNTYTVKSGDFLN-----KIAQYG-----VTV-----ANLNS 455
OY 207 WLPQFKSDARVDLNL-----RPTGGGTGYIGRNSVDMCFYDGYSTNSSLEIRFQDN 258
Db 456 W-----NGISGDLIFVGOKLIYKKGTSGMT-----GGSSNGS-----NNN 491
OY 259 PKSDGKFEYLRKINDDTKEIA--YTLS-----LLAGKSLTFPTNGTSLNIADA 303
Db 492 QSGTNTYTVKSGDILNKIAQGVSVANLRSMNGISGDLIFRQKLIYKKGTSGMTGS 551
OY 304 ASLETNMMNR 312
Db 552 SNGGSNNNQ 560

RESULT 14
AAB01830
ID AAB01830 standard; Protein; 1222 AA.
XX
AC AAB01830;
XX
DT 11-SEP-2000 (first entry)
XX
DE H. influenzae strain KI mature full-length HMW1A protein, SEQ ID NO:37.
XX
KW Mature HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW non-typeable Haemophilus influenzae; NTH1; non-encapsulated;
KW recombinant production; Escherichia coli; antibacterial; vaccine;
KW human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW detection; diagnosis.
XX
OS Haemophilus influenzae strain KI.
XX
FH Key Location/Qualifiers
FT Misc-difference 307
FT FT
FT FT
XX
PN W0200020609-A2.
XX
PD 13-APR-2000.
XX
PF 07-OCT-1999; 99WO-CA00938.
XX
PR 07-OCT-1998; 98US-0167568.
XX
PR 08-DEC-1998; 98US-0206942.
XX
PA (CONN-) CONNAUGHT LAB LTD.
XX
PI Loosmore SM, Yang Y, Klein MH;
XX
DR WPI; 2000-303789/26.
DR N-PSDB; AAA52180.
XX
PT Nucleic acid molecule for producing recombinant high molecular weight
PT proteins of Haemophilus which are used as a vaccine to provide
PT protection against Haemophilus induced diseases in humans -
XX
PS Claim 8; Fig 20A-R; 307pp; English.
XX
CC The invention relates to the recombinant production of Haemophilus
CC influenzae high molecular weight (HMW) proteins in Escherichia coli. The
CC expression construct used to effect recombinant expression comprises a
CC promoter functional in E. coli (e.g., the T7 promoter) operably linked
CC to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
CC influenzae (NTH1). Most HMW-expressing NTH1 strains contain two hmw gene
CC clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
CC hmwB and hmwC genes. The hmwA genes encode the structural HMW1A proteins
CC and the hmwB and hmwC genes encode accessory proteins which are
CC responsible for post-translational processing and secretion of the HMW1A
CC proteins. The modified hmwABC operon used in the expression construct of
CC the invention contains an A gene modified such that it encodes only the
CC mature HMW1A. The invention also discloses hmwA genes (AAA52175-A52198)

```



Sat Dec 7 16:56:46 2002

us-09-839-894-10.rag

Page 11

```

0Y 179 VK-----RRYSEYTGFTINTITLND-----KGNQJLMPQKSPARDVLDRLPTGG 227
Db 496 ISIAKGAIFRDIENSTSLNITTKSDSNHHTIKGN-----TJRKDLNITNNGN 547
0Y 228 T-YIGRNSYDMCFDYESTNSS-----LEIRFODNPKSD-GKFYLRKINDTKET 277
Db 548 TEIQIGCN---ISQKEGNLITISSQKVITERITIKAGVNDNDSNDSNATSANITTKEL 604
0Y 278 AYTSLLLAG---KSLPPTNGTSININDAASL-ETNNRITAYTMEISVPYICWGRQ 333
Db 605 KLTDMLNISGNKAEITAKDUSNLTITDSNDAGNTDARKVTFPSNVKDSKISAS--DHNYT 662
0Y 334 LDAKVE---NPEAGOVYGNINVTFPSSQTL 361
Db 663 LNSKVEITSQDTSTEDDGNNNNTGLTITAKVY 693

```

Search completed: December 4, 2002, 17:27:37  
Job time : 43 secs

**This Page Blank (uspto)**